

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 16:58:52 ; Search time 23 Seconds

(without alignments)
347,589 Million cell updates/sec

Title: US-09-171-607a-1

Perfect score: 893
Sequence: 1 VALNSPLSGMGMRGADGADQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Optimal number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	893	100.0	1516	1	CA1H_HUMAN
2	778	87.1	1527	1	CA1H_MOUSE
3	504	56.4	1388	1	CA1E_HUMAN
4	76	8.5	556	1	GUL_CHICK
5	75	8.4	271	1	Y4BG_RHISN
6	74	8.3	296	1	CRK_XENLA
7	74	8.3	780	1	SRRN_RAT
8	73.5	8.2	1653	1	YRHM_ECOLI
9	73	8.2	512	1	PPX_ECOLI
10	72.5	8.1	884	1	YP67_MYCTU
11	72	8.1	304	1	CRK_HUMAN
12	72	8.1	304	1	CRK_MOUSE
13	72	8.1	304	1	CRK_MOUSE
14	72	8.1	613	1	HS75_CANAL
15	70	7.8	780	1	STRN_MOUSE
16	70	7.8	837	1	L2TR_MOUSE
17	70	7.8	953	1	CA1E_MOUSE
18	69	7.7	266	1	CB2L_SINAI
19	69	7.7	309	1	THCC_ECOLI
20	69	7.7	326	1	THCC_ECOLI
21	69	7.7	462	1	MYCN_PSEME
22	69	7.7	1233	1	MYCN_HUMAN
23	68.5	7.7	1233	1	MYCN_HUMAN
24	68.5	7.7	1289	1	PGK_BUCAI
25	68	7.6	610	1	CSAB_BACUD
26	68	7.6	953	1	MET7_SCHPO
27	68	7.6	999	1	CAR4_HUMAN
28	67.5	7.6	999	1	MERK_HUMAN
29	67.5	7.6	348	1	NU2M_BRARE
30	67.5	7.6	536	1	CAR9_HUMAN
31	67.5	7.6	579	1	FAAH_MOUSE
32	67	7.5	692	1	GVRB_BARBA
33	67	7.5	145	1	ANF_RANCA
			250	1	HXB9_HUMAN
					P17482 homo sapien

ALIGNMENTS

RESULT 1	ID	CA1H_HUMAN	STANDARD:	PRT:	1516 AA.
AC	P39060	Q9UK38; Q9Y607; Q9Y608			
DT	01-FEB-1995	(Rel. 31, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].				
GN	COL18A1				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-98164096; PubMed-9503365;				
RA	Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;				
RT	"Complete primary structure of two variant forms of human type XVIII				
RT	collagen and tissue-specific differences in the expression of the				
RT	corresponding transcripts";				
RL	Matrix Biol. 16:319-328(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-20289799; PubMed-10830953;				
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,				
RA	Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.;				
RA	Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.;				
RA	Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.;				
RA	Reichwald K., Rump A., Schillnabel M., Schudy A., Zimmermann W.;				
RA	Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.;				
RA	Shitani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.;				
RA	Mioshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.;				
RA	Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.;				
RA	Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Dgand E.;				
RA	Weinmeyer S., Bozzym K., Gardiner K., Metzke D., Francis F.;				
RA	Leinhardt H., Reinhardt R., Yaspo M.-L.;				
RT	"The DNA sequence of human chromosome 21.;"				
RL	Nature 405:311-319(2000).				
RN	[3]				
RP	SEQUENCE OF 834-1516 FROM N.A.				
RX	MEDLINE-94245237; PubMed-8188291;				
RA	Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.;				
RA	Olsen B.R.;				
RT	"Cloning of cDNA and genomic DNA encoding human type XVIII collagen				
RT	and localization of the alpha 1(XVII) collagen gene to mouse				
RT	chromosome 10 and human chromosome 21.;"				
RL	Genomics 19:494-499(1994).				
RN	[4]				
RP	SEQUENCE OF 1334-1516 FROM N.A.				
RC	TISSUE-Placenta;				
RA	Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;				
RT	"Cloning and expression of human endostatin gene in Escherichia				
RT	coli.;"				
RT	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.				
RL	[5]				
RP	INVOLVEMENT IN KNOBLOCH SYNDROME.				

34	67	7.5	364	1	YM28_MYCTU
35	67	7.5	577	1	BAG3_MOUSE
36	66.5	7.4	368	1	GALT_HUMAN
37	66.5	7.4	449	1	DHE2_CLOSY
38	66.5	7.4	1548	1	SMCY_MOUSE
39	66	7.4	250	1	SMCY_MOUSE
40	66	7.4	342	1	G3PI_ANAVA
41	66	7.4	399	1	SUCC_CAUCR
42	66	7.4	419	1	P47K_PSECL
43	66	7.4	484	1	TRPG_YEAST
44	66	7.4	505	1	CH01_HUMAN
45	66	7.4	622	1	PPCC_MOUSE
					010512 mycobacteri
					0911v1 mus musculu
					0607v5 homo sapien
					P24295 clostridium
					062240 mus musculu
					P20615 mus musculu
					P34916 anabaena va
					09a894 caulobacter
					P31521 pseudomonas
					P00937 saccharomyc
					09y236 homo sapien
					09z2v4 mus musculu

RX MEDLINE=204000145; PubMed=109424234;
 RA Sertle A.L., Sossi V., Camargo A.A., Zatz M., Brane C.,
 RA Passos-Bueno M.R.,
 RT "Collagen XVII, containing an endogenous inhibitor of angiogenesis
 RT and tumor growth, plays a critical role in the maintenance of retinal
 RT structure and in neural tube closure."
 RL Hum. Mol. Genet. 9:2051-2058(2000).
 RN [6]
 RP VARIANT ASN-1437.
 RX MEDLINE=21518361; PubMed=11606364;
 RA Iuguet P., Suzuki O., Godol P.H., Alves V.A., Sertle A.L.,
 RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
 RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.,
 RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
 RT for the development of prostatic adenocarcinoma."
 RL Cancer Res. 61:7375-7378(2001).
 CC -1- FUNCTION: COLA1A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
 CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
 CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
 CC FACTOR SIGNALING.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing: Named isoforms=2;
 CC Name-Long: Synonyms=NC-493;
 CC IsoId=P39060-1; Sequence=Displayed;
 CC Name-Short: Synonyms=NC1-303;
 CC IsoId=P39060-2; Sequence=VSP_001155, VSP_001156;
 CC -1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
 CC IN LIVER, LUNG AND KIDNEY.
 CC -1- PM: Prolins are the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- POLYMORPHISM: There is an association between a polymorphism in
 CC position 1437 and prostate cancer. Heterozygous Asn-1437
 CC individuals have a 2.5 times increased chance of developing
 CC prostate cancer as compared with homozygous Asp-1437 individuals.
 CC -1- DISEASE: Defects in COL1A1 are a cause of Knobloch syndrome (KNO)
 CC [MIM:267750]; an autosomal recessive disorder defined by the
 CC occurrence of high myopia, vitreoretinal degeneration with retinal
 CC detachment, macular abnormalities and occipital encephalocele.
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage for and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF018081; AAC39658.1; -
 DR EMBL; AF018082; AAC39659.1; -
 DR EMBL; AL163302; CAB90482.1; -
 DR EMBL; L22548; AAS1864.1; -
 DR EMBL; AF184060; AAF01310.1; ALT_INIT.
 DR PDB; 1BNL; O2-DEC-98.
 DR GLYcosultEDB; P39060; -
 DR GeneW; HGNC:2195; COL18A1.
 DR MIM; 120328; -
 DR MIM; 267750; -
 DR GO; GO:0005581; C:collagen; TAS.
 DR GO; GO:0008181; F:tumor suppressor; TAS.
 DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.
 DR GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
 DR GO; GO:0007048; P:oncogenesis; TAS.
 DR GO; GO:0007601; P:vision; TAS.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 7.
 DR Pfam; PF02210; TSPN; 1.
 DR ProDom; PD000007; Ctg_helix; 1.

[illegible]


```

FT DOMAIN 674 795 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 796 818 NONHELIICAL REGION 4 (NC4).
FT DOMAIN 819 901 TRIPLE-HELICAL REGION 4 (COL4).
FT DOMAIN 902 915 NONHELIICAL REGION 5 (NC5).
FT DOMAIN 916 957 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 958 970 NONHELIICAL REGION 6 (NC6).
FT DOMAIN 971 1043 TRIPLE-HELICAL REGION 6 (COL6).
FT DOMAIN 1044 1053 NONHELIICAL REGION 7 (NC7).
FT DOMAIN 1054 1086 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1087 1098 NONHELIICAL REGION 8 (NC8).
FT DOMAIN 1099 1122 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1123 1132 NONHELIICAL REGION 9 (NC9).
FT DOMAIN 1130 1181 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1182 1194 NONHELIICAL REGION 10 (NC10).
FT DOMAIN 1195 1212 TRIPLE-HELICAL REGION 10 (COL10).
FT DOMAIN 1213 1527 NONHELIICAL REGION 11 (NC11).
FT CARBOHYD 338 338 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 1376 1516 MISSING (in isoform Short).
FT DISULFID 1478 1508 MISSING (in isoform Short).
FT SITE 1104 1106 MISSING (in isoform Short).
FT VARSPLIC 1 212
FT VARSPLIC 213 238
FT CONFLICT 900 900
FT CONFLICT 947 947 P -> L (IN REF. 4).
FT CONFLICT 964 964 A -> R (IN REF. 4).
FT CONFLICT 1157 1157 R -> P (IN REF. 4).
FT CONFLICT 1266 1266 P -> L (IN REF. 4).
FT CONFLICT 1276 1276 L -> F (IN REF. 4).
FT CONFLICT 1437 1437 L -> V (IN REF. 4).
FT STRAND 1353 1357
FT STRAND 1362 1362
FT STRAND 1368 1382
FT STRAND 1383 1384
FT STRAND 1389 1391
FT STRAND 1395 1395
FT STRAND 1396 1397
FT STRAND 1398 1398
FT STRAND 1403 1403
FT HELIX 1406 1408
FT TURN 1409 1411
FT STRAND 1411 1415
FT TURN 1414 1415
FT STRAND 1417 1418
FT STRAND 1421 1423
FT HELIX 1426 1429
FT TURN 1430 1430
FT TURN 1436 1436
FT TURN 1439 1440
FT STRAND 1444 1445
FT STRAND 1446 1447
FT TURN 1450 1450
FT TURN 1451 1453
FT STRAND 1455 1456
FT STRAND 1461 1463
FT STRAND 1466 1466
FT TURN 1468 1469
FT STRAND 1472 1472
FT TURN 1474 1475
FT STRAND 1477 1477
FT HELIX 1478 1481
FT TURN 1482 1482
FT TURN 1486 1487
FT STRAND 1489 1494
FT HELIX 1495 1497

Query Match 87.1%; Score 778; DB 1; Length 1527;
Best Local Similarity 85.8%; Pred. No. 3-3e-69;
Matches 145; Conservative 13; Mismatches 11; Indels 0; Gaps 0;
j VALNSPLSGMGINGADFCQOARAVAGLAGTFRAFLSSRLQDLYSIVRRADRAAVIV 60

```

```

Db 1356 VALNPLPSGGMGINGADFCQOARAVAGLAGTFRAFLSSRLQDLYSIVRRADRAAVIV 1415
Qy 61 NIKDELLPSWALSGSESGPLKPGARIFSPFGKDVLRHPWPQKSVHMGSDPNRRRLTE 120
Db 1416 NIKDEVLSWSDLSGSGGQIQPGARITSPFGKDVLRHPWPQKSVHMGSDPNRRRLME 1475
Qy 121 SYCETWRTPAPSATGQASSLLGGRLLGQSSAACHAYIVLCIENSFMTA 169
Db 1476 SYCETWRTPETGATGQASSLLGGRLLGQSSAACHNSYIVLCIENSFMTS 1524

RESULT 3
CAIE_HUMAN STANDARD; PRT; 1388 AA.
AC P39059;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XV) chain precursor.
GN COL15A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical cord;
RX MEDLINE=94148920; PubMed=8106446;
RA Kivirikko S., Helmanaki P., Rehn M.V., Honkanen N., Myers J.C.,
RA Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of human type XV collagen and
RT exon-intron organization in the 3' region of the corresponding
RT gene.";
RL J. Biol. Chem. 269:4773-4779(1994).
RN [2]
RP SEQUENCE OF 1-569 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94140817; PubMed=8307960;
RA Muragaki Y., Abe N., Niinomiya Y., Olsen B.R., Ooshima A.;
RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
RT non-triple helical domain with a tandem repeat structure and homology
RT to alpha 1(XVII) collagen.";
RL J. Biol. Chem. 269:4042-4046(1994).
RN [3]
RP SEQUENCE OF 544-1252 FROM N.A.
RX MEDLINE=93066196; PubMed=1279671;
RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
RT "Identification of a previously unknown human collagen chain, alpha
RT 1(XV), characterized by extensive interruptions in the triple-helical
RT region.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
RN [4]
RP TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
CC SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL, L25286; AAA58429.1; -
DR EMBL, D21230; BAA04762.1; -
DR EMBL, L01697; -; NOT_ANNOTATED_CDS.
DR PIR, A53317; A53317.
DR HSSP, P39061; IKOE.
DR Genew; HGNC:2192; COL15A1.

```

```

DR MIM: 120325;
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001791; Lamln1_g.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01391; Collagen; 5.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamC; 1.
DR SMART: SM00210; TSPN; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1388
FT DOMAIN 26 555
FT DOMAIN 556 573
FT DOMAIN 574 618
FT DOMAIN 619 732
FT DOMAIN 733 763
FT DOMAIN 764 798
FT DOMAIN 799 822
FT DOMAIN 823 867
FT DOMAIN 868 878
FT DOMAIN 879 949
FT DOMAIN 950 983
FT DOMAIN 984 1013
FT DOMAIN 1014 1027
FT DOMAIN 1028 1045
FT DOMAIN 1046 1052
FT DOMAIN 1053 1107
FT DOMAIN 1108 1117
FT DOMAIN 1118 1132
FT DOMAIN 1133 1388
FT REPEAT 358 408
FT REPEAT 409 459
FT REPEAT 460 509
FT REPEAT 510 555
FT CARBOHYD 306 306
FT CARBOHYD 324 324
FT CARBOHYD 687 687
FT CARBOHYD 807 807
FT CARBOHYD 814 814
FT CARBOHYD 1046 1046
FT CONFLICT 10 10
FT CONFLICT 49 49
FT CONFLICT 95 95
FT CONFLICT 150 150
FT CONFLICT 204 204
FT CONFLICT 409 409
FT SEQUENCE 1388 AA: 141930 MW: 60822AD925A3093D CRC64;

Query Match
Best Local Similarity 56.4%; Score 504; DB 1; Length 1388;
Matches 95; Conservative 27; Mismatches 41; Indels 4; Gaps 1;

QY 2 ALNSPLSGMKGIRGADFOCFQOARAVAGTFRPFTSSRLDLYSIYRRADRAVPIVN 61
DB 1222 ALNMFSGDIR---ADFCQKQARAAGLSTYRAFLSHQDLSTYRKAKRISLPIVN 1277
QY 62 LKDELLEPSEWALEFSGSEGLPKRGARLFSDGKDVLRHPWPQKSVHGSOPNGRRLTES 121
DB 1278 LKGOVLENNWDSITSGGOFNMHPIYSPGRDIMDPQKVIWHGSSPHGVRLVDN 1337
QY 122 YCETWTFEAPSATGOASSLIGRRLGSSASCHNAVYLCIENEFMT 168
DB 1338 YCEAWRTRADYAVTGCLASPLSTGKILDKAKYSCANRLVLCIENSEMT 1384

RESULT 4
ID GLI_CHICK STANDARD; PRT; 556 AA.
AC P55878;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

```

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein GLI1 (GLI) (Fragment).
GN GLI OR GLI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97105842; PubMed=8948590;
RA Marigo V., Johnson R.L., Vortkamp A., Tabin C.J.;
RT "Sonic hedgehog differentially regulates expression of GLI and GLI3
RL during limb development.";
CC Dev. Biol. 180:273-283(1996).
CC -I- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
CC NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
CC AND DIGITAL DEVELOPMENT. AS WELL AS DEVELOPMENT OF THE CENTRAL
CC NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
CC TRANSDUCTION OF SHH SIGNAL. (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
DR EMBL: U60762; AAB51659.1; -.
DR HSP: P08151; 26LI.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00096; Zf_C2H2; 5.
DR SMART: PF00355; Znf_C2H2; 5.
DR PROSITE: PS00287; ZINC_FINGER_C2H2_1; 4.
DR PROSITE: PS00577; ZINC_FINGER_C2H2_2; 5.
KW Zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
KW Nuclear protein; Repeat.
FT ZN_FING 247 272 C2H2-TYPE.
FT ZN_FING 280 307 C2H2-TYPE.
FT ZN_FING 313 337 C2H2-TYPE.
FT ZN_FING 343 368 C2H2-TYPE.
FT ZN_FING 374 399 C2H2-TYPE.
FT NON_TER 556
SQ SEQUENCE 556 AA: 60215 MW: 722D2AA5A1CA4D98 CRC64;

Query Match
Best Local Similarity 8.5%; Score 76; DB 1; Length 556;
Matches 48; Conservative 21; Mismatches 66; Indels 72; Gaps 10;

QY 6 PLSG---GMGRIGRADRPQ-C-----PQOARAVGLAG 32
DB 20 PLHSAAGTPELOGLDLPVCHQPMVASSHNGYGLVPTENHGGAADSRFTSPGAGKLG 79
QY 33 TFRAF---LSSRLDLYSIYRRADRAVPIVN-----LKDELLEPS----- 70
DB 80 KKRALSLSPSSDSSVDLOTIVRTSPNSIVAFINRSCASAGSGYHLSITSPSLGYQNP 139
QY 71 -----WEALFSGSEGLPKRGARLFSDGKDVLRHPWPQKSVHGSOPNGRR-----LT 119
DB 140 PGQOGGOGOLFSTHP-PLPPCSSHETLSRGLHPTPARCTIKHCQOLKERSLSPLT 198
QY 120 ESYCETWTFE---APSATGOASSLIG 142
DB 199 AKIPEE-KSEGDISSPASTGTODPLLG 224

RESULT 5
ID Y4BG_RHISN STANDARD; PRT; 271 AA.
DT Y4BG_RHISN

```

```

AC P55374.1 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical protein Y4BG precursor.
GN Y4BG.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym PNGR234e.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RL Nature 387:394-401(1997).
-1 SIMILARITY: NONE OBTAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sdb.ch/announce/
or send an email to license@sdb-sdb.ch).
-----
DR EMBL; AE000066; AAB91622.1; -
KW Hypothetical protein; Plasmid; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 271 HYPOTHEMETICAL PROTEIN Y4BG.
SQ SEQUENCE 271 AA: 30034 MW: 5572CF1461FCD463 CRC64;

Query Match 8.4%; Score 75; DB 1; Length 271;
Best Local Similarity 21.1%; Pred. No. 3.5;
Matches 28; Conservative 26; Mismatches 35; Indels 44; Gaps 7;

OY 4 NSPLSGGKRGIRGADFCFOQARAVGLAGTFRALSLRLOLYSVRRA-----DRAAV 57
DB 174 NADIAATRTKSLDGLADFEAFQAIITGSGSEFKS-----DDY-IGRCPTPHMGREORAL 225
OY 58 PIVNLKDELDELPSWEALFSGSEGPLKPARIRFSPGKDYLNHP--TWPOKSYVHGSOPN 114
DB 226 LFLSAKDDRAVAAW-----KP-----HOKKIIVPYKQWPKA----- 259
OY 115 GRRLTESYCEYTW 127
DB 260 -KOELRAMAETWK 271

RESULT 6
CRK_XENLA STANDARD; PRT; 296 AA.
ID CRK_XENLA P87378;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SH2/SH3 adaptor crk (Adapter molecule crk) (CRK2).
GN CRK.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97180926; PubMed=9029144;
RA Evans E.K., Lu W., Strum S.L., Meyer B.J., Kornbluth S.;
RL "Crk is required for apoptosis in Xenopus egg extracts."
EMBO J. 16:230-241(1997).
-1- FUNCTION: Required for apoptosis in Xenopus egg extracts. May
mediate attachment-induced MAPK8 activation, membrane ruffling and

```

```

CC      cell motility in a Rac-dependent manner.
CC      -1 SUBUNIT: Interacts with ABL, C3G, SOS, MAP4K1, MAPK8 and DOCK1 via
CC      its first SH3 domain. Interacts with BCAR1, CBL, PXN and GAB1 via
CC      its SH2 domain upon stimulus-induced tyrosine phosphorylation.
CC      Interacts with several tyrosine-phosphorylated growth factor
CC      receptors such as EGFR, PDGFR and INSR via its SH2 domain (By
CC      similarity).
CC      -1 SUBCELLULAR LOCATION: Cytoplasmic; translocated to the plasma
CC      membrane upon cell adhesion (By similarity).
CC      -1 PTM: Phosphorylated on Tyr-214 upon cell adhesion. Results in the
CC      negative regulation of the association with SH2- and SH3-binding
CC      partners, possibly by the formation of an intramolecular
CC      interaction of phosphorylated Tyr-221 with the SH2 domain. This
CC      leads finally to the down-regulation of the Cdk signaling pathway
CC      (By similarity).
CC      -1 SIMILARITY: Contains 1 SH2 domain.
CC      -1 SIMILARITY: Contains 2 SH3 domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      DR EMBL; 089774; ABA49698.1; -.
CC      DR HSSP; Q64010; ICKA.
CC      DR InterPro; IPR000980; SH2.
CC      DR InterPro; IPR001452; SH3.
CC      DR Pfam; PF000017; SH2; 1.
CC      DR Pfam; PF000018; SH3; 1.
CC      DR PRINTS; PR00401; SH2DOMAIN.
CC      DR PRINTS; PR00452; SH3DOMAIN.
CC      DR ProDom; PD000093; SH2; 1.
CC      DR ProDom; PD000066; SH3; 1.
CC      DR SMART; SM00252; SH2; 1.
CC      DR SMART; SM00326; SH3; 1.
CC      DR PROSITE; PS50001; SH2; 1.
CC      DR PROSITE; PS50002; SH3; 1.
CC      KW SH2 domain; SH3 domain; Repeat; Phosphorylation.
CC      FT DOMAIN 13 112 SH2.
CC      FT DOMAIN 125 185 SH3 1.
CC      FT DOMAIN 249 289 SH3 2.
CC      FT MOD_RES 214 214 PHOSPHORYLATION (By SIMILARITY).
CC      SQ SEQUENCE 296 AA; 33409 MM; 544FLIFAI1F75A66 CRC64;
CC      -----
CC      Query Match 8.3%; Score 74; DB 1; Length 296;
CC      Best Local Similarity 32.4%; Pred. NO. 4.9;
CC      Matches 24; Conservative 12; Mismatches 20; Indels 18; Gaps 5;
CC      -----
QY 75 FSGSEG---PLKPGARIFSGDKDVLRLPPTPKQSVHGSNDPNGRR--LTESYCEWTE 129
QY 136 FIGNDDDELPPKKG-----DILRIREKPEQGMNNAEDSDGRGMIPPYEKIR-- 184
QY 130 APSATGQASSILGG 143
QY 11: 11:111
Db 185 PPSSPG--SALIGG 196

RESULT 7
STRN_RAT STANDARD; PRT; 780 AA.
AC P70483;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Striatin.
GN STRN.
OS Rattus norvegicus (Rat).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;

```


OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OC NCBI_TaxID=562, 217992, 83334;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
 RC STRAIN-K12; Pubmed-8380170;
 RA MEDLINE-93107072; Pubmed-8380170;
 RA Akiyama M., Crooke E., Kornberg A.;
 RT "An exopolyphosphatase of *Escherichia coli*. The enzyme and its ppx
 RT gene in a polyphosphate operon."
 RT J. Biol. Chem. 268:633-639(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426517; Pubmed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of *Escherichia coli* K-12."
 RA Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97349980; Pubmed-9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiiuchi T.;
 RA "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features."
 RT DNA Res. 4:91-113(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE-22388334; Pubmed-12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Raso D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RA "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic *Escherichia coli*."
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE-21074935; Pubmed-11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamocis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7."
 RT Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE-21156231; Pubmed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12."
 RT DNA Res. 8:11-22(2001).
 RN [7]
 RP SIMILARITY TO GPPA.
 RX MEDLINE-94025037; Pubmed-8212131;
 RA Reitzer J., Reitzer A., Sailer M.H. Jr., Bork B., Sander C.;

RT "Exopolyphosphate phosphatase and guanosine pentaphosphate
 RT phosphatase belong to the sugar kinase/actin/hsp 70 superfamily.";
 RT Trends Biochem. Sci. 18:247-248(1993).
 CC -1- FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE
 CC IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA
 CC 500 RESIDUES LONG, BUT CHAINS OF CIRCA 15 RESIDUES COMPLETE POORLY
 CC WITH POLYPHOSPHATE AS SUBSTRATE.
 CC -1- CATALYTIC ACTIVITY: (polyphosphate)(N) + H(2)O =
 CC (polyphosphate)(N-1) + phosphate.
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.
 CC -1- SIMILARITY: BELONGS TO THE GPPA / Ppx FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 DR EMBL: L06129; AAA24415.1; -
 DR EMBL: AE000336; AAC75555.1; -
 DR EMBL: D90878; BAA16390.1; -
 DR EMBL: D90880; BAA16392.1; -
 DR EMBL: AE016764; AAN81470.1; -
 DR EMBL: AE005479; AAG57612.1; -
 DR EMBL: AP002561; BAB36787.1; -
 DR PIR: A45333; A45333.
 DR PIR: D91049; D91049.
 DR PIR: H85893; H85893.
 DR Ecogene: EG11403; ppx.
 DR InterPro: IPR003695; Ppx_Gppa.
 DR Pfam: PF02541; Ppx-Gppa.1.
 DR HydroLase; Magnesium; Membrane; Complete proteome.
 FT INIT_MET 0
 FT MET 0
 SQ SEQUENCE 512 AA; 58004 MW; 48611AFF5D9FB9C3 CRC64;
 Query Match 8.2%; Score 73; DB 1; Length 512;
 Best Local Similarity 21.9%; Pred. No. 12;
 Matches 46; Conservative 28; Mismatches 66; Indels 70; Gaps 11;
 QY 22 PQQAR-----AVGLAQTFR-----FLSSRLQDLY 46
 DB 186 PQRAMAAQAQKLETTWQFRIGMNVAVAGASTTIAAEHVLMEKNGKIIITPERLEKLY 245
 QY 47 -SVRRADRAVPIYNLKE--LFPSSWELFSG-----SEGLPKGARIFSF 91
 DB 246 KEVLRRHFASLSLPGLESEKRTVPGLAILGVFDALAIRELRLSGALREGV-LYEM 304
 QY 92 DKG---DVLRRPPTPKQSVHMGSDPNGRRL---TESYCTWRTEAPS-ATGOASSLLG- 142
 DB 305 EGRFHQVRSRTASSLANQYHIIDSEQARRVLDITTMQMEQVREQQPLAHQDEALLRW 364
 QY 143 -----GRLLGQSAASCHHAYIVLCIENS 165
 DB 365 AAMLHEVGLNINHSGLHNSAYI---LQNS 391
 RESULT 10
 YP67 MYCTU
 ID YP67 MYCTU STANDARD; PRT; 884 AA.
 AC OS0654; OS0731;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 30-MAY-2000 (rel. 39, Last sequence update)
 DT 28-FEB-2003 (rel. 41, Last annotation update)
 DE Hypothetical protein RV2567.
 GN RV2567 OR MT2643 OR MTCY227.34C OR MTCY9C4.01C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;

CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic Lung, and Placenta;
RX MEDLINE=92334347; PubMed=1630456;
RA Matsuda M., Tanaka S., Nagata S., Kojima A., Kurata T., Shibuya M.;
RT "Two species of human Crk cDNA encode proteins with distinct
RL biological activities."; Mol. Cell. Biol. 12:3482-3489(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93390962; PubMed=8378094;
RA Fiorello T., Helsterkamp N., Groffen J., Benjes S., Morris C.;
RT "CRK proto-oncogene maps to human chromosome band 1P13.";
RL Oncogene 8:2853-2855(1993).
RN [3]
RP INTERACTION WITH DOCK1.
RX MEDLINE=96239533; PubMed=8657152;
RA Hasegawa H., Kiyokawa E., Tanaka S., Nagashima K., Gotch N.,
RA Shibuya M., Kurata T., Matsuda M.;
RT "DOCK180, a major Crk-binding protein, alters cell morphology upon
translocation to the cell membrane.";
RL Mol. Cell. Biol. 16:1770-1776(1996).
RN [4]
RP INTERACTION WITH DOCK1; CSG AND EPS15, AND MUTAGENESIS OF
ASP-150.
RX PubMed=8662907;
RA Matsuda M., Ota S., Tanimura R., Nakamura H., Matuoka K., Takenawa T.,
RA Nagashima K., Kurata T.;
RT "Interaction between the amino-terminal SH3 domain of Crk and its
natural target proteins.";
RL J. Biol. Chem. 271:14468-14472(1996).
RN [5]
RP INTERACTION WITH DOCK4.
RX MEDLINE=22515523; PubMed=12628187;
RA Tajiki V., Pauling G., Sordele A.R., McClatchey A.I., Saeto M.,
RA Wainwright D.C.R., Reynolds P., Bell D.W., Lake R., van den Heuvel S.,
RA Settleman J., Haber D.A.;
RT "Dock4, a Grap2 activator, is disrupted during tumorigenesis.";
RL Cell 112:673-684(2003).
CC -I- FUNCTION: The Crk-I and Crk-II forms differ in their biological
activities. Crk-II has less transforming activity than Crk-I. Crk
II mediates attachment-induced MAPK activation, membrane ruffling,
and cell motility in a Rac-dependent manner. Involved in
phagocytosis of apoptotic cells and cell motility via its
interaction with Dock1 and Dock4.
CC -I- SUBUNIT: Interacts with Abl, Csg, Sos, Mapk1, Mapk8 and Dock3 via
its first SH3 domain. Interacts with BcrA1, Cbl, Pxn and Gab1 via
its SH2 domain upon stimulus-induced tyrosine phosphorylation.
Interacts with several tyrosine-phosphorylated growth factor
receptors such as EGFR, PDGFR and INSR via its SH2 domain (By
similarity). Interacts with Dock1 and Dock4.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic; translocated to the plasma
membrane upon cell adhesion (By similarity).
CC -I- ALTERNATIVE PRODUCTS:
CC Name-Crk-I;
CC Name-Crk-II;
CC Event=Alternative splicing; Named isoforms=2;
CC IsoId=P46108-1; Sequence=Displayed;
CC Name-Crk-I;
CC IsoId=P46108-2; Sequence=VSP_004173;
CC -I- DOMAIN: The C-terminal SH3 domain function as a negative modulator
for transformation and the N-terminal SH3 domain appears to
function as a positive regulator for transformation (By
similarity).
CC -I- PTM: Phosphorylation of Crk-II (40 kDa) gives rise to a 42 kDa
form.
CC -I- PTM: Phosphorylated on Tyr-221 upon cell adhesion. Results in the
negative regulation of the association with SH2- and SH3-binding
partners, possibly by the formation of an intramolecular
interaction of phosphorylated Tyr-221 with the SH2 domain. This
leads finally to the down-regulation of the Crk signaling pathway
(By similarity).


```

CC EMBL: S72408; AAB30755.1;
DR PDB: 1CKA; 08-MAY-95.
DR PDB: 1CKB; 08-MAY-95.
DR PDB: 1B07; 06-JAN-99.
DR PDB: 1J05; 06-NOV-02.
DR MGI: 88508; Crk.
DR GO: GO:0005515; P:protein binding activity; IPI.
DR InterPro: IPR000980; SH2.
DR Pfam: PF00017; SH2; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH2DOMAIN.
DR ProDom: PD000093; SH2; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;
Phosphorylation; 3D-structure.
FT DOMAIN 13 118 SH2.
FT DOMAIN 132 192 SH3 1.
FT DOMAIN 256 296 SH3 2.
FT MOD_RES 221 221 PHOSPHORYLATION (BY SIMILARITY).
FT VARSPLIC 205 304 Missing (in isoform Crk-I).
/FTID=VSP_004174.

FT STRAND 136 139
FT STRAND 143 143
FT TURN 148 149
FT STRAND 150 150
FT STRAND 153 153
FT TURN 155 156
FT STRAND 158 163
FT STRAND 169 173
FT TURN 175 176
FT STRAND 179 183
FT HELIX 184 186
FT STRAND 187 188
SO SEQUENCE 304 AA; 33814 MW; 5491896FC7A89065 CRC64;

Query Match 8.1%; Score 72; DB 1; Length 304;
Best Local Similarity 27.8%; Pred. NO. 8;
Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

QY 88 IFSEFGK-----DVLHPTWPKSVHSGSDPNRR--LRESYETWRTSPATG 135
140 LFDPNGNDEEDLPKKGDILIRKPKPEQWMAEDSEGRGMIPVPEYKTR---PASA 195
136 QASSLLGGRLLG 147
196 SVSALLGNGEG 207

Db
196 SVSALLGNGEG 207

RESULT 13
CRK_RAT
ID CRK_RAT STANDARD: PRT: 304 AA.
AC 063768;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Proto-oncogene C-crk (P38) (adapter molecule crk).
GN CRK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97057214; PubMed=8901553;
RA Kizaka-Kondoh S., Matsuda M., Okayama H.;

```

```

RT "CrkII signals from epidermal growth factor receptor to Ras."
RL Proc. Natl. Acad. Sci. U.S.A. 93:12177-12182(1996).
RN (2)
RP PHOSPHORYLATION OF TYR-221, AND MUTAGENESIS OF TYR-221.
RX MEDLINE=22186593; PubMed=12198159;
RA Abassi Y.A., Vuori K.;
RT "Tyrosine 221 in Crk regulates adhesion-dependent membrane
RT localization of Crk and Rac and Rac activation of Rac signaling."
RL EMBO J. 21:4571-4582(2002).
CC -1- FUNCTION: The Crk-I and Crk-II forms differ in their biological
CC activities. Crk-II has less transforming activity than Crk-I. Crk-
CC II mediates attachment-induced MAPK8 activation, membrane ruffling
CC and cell motility in a Rac-dependent manner.
CC -1- SUBUNIT: Interacts with DOCK1, DOCK3, DOCK4, C3G and EPS15 via its
CC SH3 domain (by similarity). Interacts with ABL, C3G, SOS, MAPK1,
CC MAPK8 and DOCK1 via its first SH3 domain. Interacts with BCRKL,
CC CBL, FXN and GAB1 via its SH2 domain upon stimulus-induced
CC tyrosine phosphorylation. Interacts with several tyrosine-
CC phosphorylated growth factor receptors such as EGFR, PDGFR and
CC INSR via its SH2 domain.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; translocated to the plasma
CC membrane upon cell adhesion.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=Crk-II;
CC IsoId=Q63768-1; Sequence=Displayed;
CC Name=Crk-I;
CC IsoId=Q63768-2; Sequence=VSP_004175;
CC -1- TISSUE SPECIFICITY: Crk-II is expressed in all tissues and cells
CC whereas Crk-I is expressed at lower level and in limited cell-
CC types.
CC -1- DOMAIN: The C-terminal SH3 domain function as a negative modulator
CC for transformation and the N-terminal SH3 domain appears to
CC function as a positive regulator for transformation (by
CC similarity).
CC -1- PTM: Phosphorylated on Tyr-221 upon cell adhesion. Results in the
CC negative regulation of the association with SH2- and SH3-binding
CC partners, possibly by the formation of an intramolecular
CC interaction of phosphorylated Tyr-221 with the SH2 domain. This
CC leads finally to the down-regulation of the Crk signaling pathway.
CC -1- SIMILARITY: Contains 1 SH2 domain.
CC -1- SIMILARITY: Contains 2 SH3 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D44481; BAA07924.1;
DR HSSP: Q64010; 1CKA.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000093; SH2; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00252; SH2; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
DR Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;
KW Phosphorylation.
FT DOMAIN 13 118 SH2.
FT DOMAIN 132 192 SH3 1.
FT DOMAIN 256 296 SH3 2.
FT MOD_RES 221 221 PHOSPHORYLATION.
FT VARSPLIC 205 304 Missing (in isoform Crk-I).
/FTID=VSP_004175.

```

FT VARIANT 244 244 Q -> R (NRK-23 INACTIVE MUTANT).
 FT VARIANT 253 253 K -> E (NRK-23 INACTIVE MUTANT).
 FT MUTAGEN 221 221 Y -> F: NO ACTIVATION OR RAC SIGNALING.
 SQ SEQUENCE 304 AA: 33844 MW: 4CBBP65BE722265 CRC64;

Query Match 8.1%; Score 72; DB 1; Length 304;
 Best Local Similarity 27.8%; Pred. No. 8;

Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

QY IFSPECK-----DVLRHPTWPKSVHSGSPNRR--LTSTCYETWTEPATSATG 135
 DB 140 LFDENGNDDEDLPEKKGILIRDKPEBOMWNAEDSEKRGMPVYVEKYR-----PASA 195

QY 136 QASSILGRLG 147
 DB 196 SVSALIGNGEG 207

RESULT 14

75-CANAL
 HS75-CANAL STANDARD; PRT: 613 AA.

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Heat shock protein SSB1.
 GN SSB1 OR HSP70B.
 OS Candida albicans (Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
 OX NCBI_Taxid=3476;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MO-1;
 RX MEDLINE=97344370; PubMed=9200817;
 RA Maneu V., Cervera A.M., Martinez J.P., Gozalbo D.:
 RT Molecular cloning of a Candida albicans gene (SSB1) coding for a
 RL protein related to the Hsp70 family.";
 Yest 13:677-681(1997).

-1- FUNCTION: MAY AID IN THE PASSAGE OF THE NASCENT POLYPEPTIDE CHAIN
 THROUGH THE RIBOSOME CHANNEL INTO THE CYTOSOL. SUCH AN INTERACTION
 COULD BE CRUCIAL FOR CONTINUOUS TRANSPORT OF THE POLYPEPTIDE;
 WITH TRANSLATION BY CLOGGING THE RIBOSOME CHANNEL (BY SIMILARITY).
 CC SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH TRANSLATING
 CC RIBOSOMES; MAY BIND DIRECTLY TO THE NASCENT POLYPEPTIDE (BY
 CC SIMILARITY).

-1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X97723; CAAG6308.1;
 DR HSP: P08107; 1HJ0.
 DR COMPLUYEAST-2DPAGE; P87222;
 DR InterPro: IPR001023; Hsp70.
 DR Pfam: PF00012; HSP70.1.
 DR PRINTS: PR00301; HEATSHOCK70.
 DR PRODOM: PD000089; HSP70.1.
 DR PROSITE: PS00297; HSP70_1.1.
 DR PROSITE: PS00329; HSP70_2.1.
 DR PROSITE: PS01036; HSP70_3.1.
 KW Heat shock; ATP-binding; Multigene family; Protein biosynthesis.
 SQ SEQUENCE 613 AA: 66432 MW: 5FFBA9F8F9327F9 CRC64;

Query Match 8.1%; Score 72; DB 1; Length 613;
 Best Local Similarity 25.6%; Pred. No. 18;
 Matches 33; Conservative 20; Mismatches 50; Indels 26; Gaps 6;

QY 32 GTFRAFLSRLQDLSYVRADRAAVPIYNLKDELLPSPWEALFS-----GSEGPL 82
 DB 4 GVFQAGIGIDLTSTTCVATYDIAVEILLANEGGNRVTSEVAFTEERLIGDAKNAQAL 63

QY 83 KFGARIF-----SPDGKDLRH-PTWPKSVHSGSPNRR-LTE-SYCTWPTAPS 132
 DB 64 NPKNTVFPAAKRLRIGAFDESVQKDKISMPFRVY-----ESNQPLIEVYIDETKTFSPQ 119

QY 133 ATGQASSLL 141
 DB 120 ---EISSMV 125

RESULT 15

STRN_MOUSE STANDARD; PRT: 780 AA.

AC 055106;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Striatin.
 GN STRN.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RA Mogrich A., Mattei M.-G., Bartoli M., Rakitina T., Baillat G.,
 RA Monneron A., Castets F.;
 RL Submitted (Jan-1998) to the EMBL/GenBank/DBJ databases.

[2]
 TISSUE SPECIFICITY.
 RX MEDLINE=20347911; PubMed=10748158;
 RA Castets F., Rakitina T., Galliard S., Mogrich A., Mattei M.-G.,
 RA Monneron A.;
 RT "Znneonin, SGNA, and striatin are calmodulin-binding, WD repeat
 RT proteins principally expressed in the brain.";
 RL J. Biol. Chem. 275:19970-19977(2000).

-1- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER. MAY
 CC FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.
 CC -1- SUBUNIT: INTERACTS WITH PROTEIN PHOSPHATASE 2A (PP2A) (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND.
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN BRAIN BUT IS ALSO
 CC EXPRESSED AT LOW LEVELS IN VARIOUS TISSUES SUCH AS KIDNEY, SPLEEN,
 CC SKELETAL MUSCLE AND LUNG.
 CC -1- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.
 CC -1- SIMILARITY: Contains 6 WD repeats.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ223777; CAAL1545.1;
 DR MGD: MG1:1333757; Strn.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40.6
 DR PRINTS: PR00320; GPROTEINRPT.
 DR PRODOM: PD000018; WD40.2.
 DR SMART: SM00320; WD40.6.
 DR PROSITE: PS00678; WD_REPEATS_1; 3.
 DR PROSITE: PS00682; WD_REPEATS_2; 4.
 DR PROSITE: PS50294; WD_REPEATS_REGION.1.
 KW Calmodulin-binding; Repeat; WD repeat; Coiled coil.
 KW DOMAIN 53 120 COILED COIL (POTENTIAL).
 FT DOMAIN 149 166 CALMODULIN-BINDING (POTENTIAL).
 FT REPEAT 461 500 WD 1.

FT REPEAT 514 553 WD 2.
FT REPEAT 567 606 WD 3.
FT REPEAT 662 701 WD 4.
FT REPEAT 704 743 WD 5.
FT REPEAT 750 779 WD 6.
FT DOMAIN 37 45 POLY-ALA.
FT SITE 55 63 CAVEOLIN-BINDING (POTENTIAL).
SQ SEQUENCE 780 AA: 86013 MW; DBD1104FF9E5BC08 CRC64;

Query Match 7.8%; Score 70; DB 1; Length 780;
Best Local Similarity 23.3%; Pred. No. 38;
Matches 50; Conservative 19; Mismatches 66; Indels 80; Gaps 12;

QY 2 ALNSPLSGMGRIGADFCQFOQARAAG-LAGTFRAFLSSRLQDLYSIVRADRAAVPIY 60
DB 398 ALTFPPSGSGKSFIMGAD-EALESELGLGELAG-----LTVANEADSLAYDIA 443
QY 61 NIKDELLFPSPWEALFSGSEGLPKPGARIFSPDGKDYLR-HPTWP----- 103
DB 444 NNKDALR-KTWNPKEF-----LRS-----HFDGIRALAFHPPIEPVLTASEDHITLKMNL 492
QY 104 -----QKSVMHGSDP-----NGRRL-----TESYCETWRTAPSA-- 133
DB 493 OKTAPAKKSTSDVEPIYTFRAHKGPLYLCVWSSNGEQCYSGGTDGRIQSWSTPNPNDP 552
QY 134 -TGQASSLLGRRLLGOS-----AASCHAYIVLC 161
DB 553 YDAYDPVYLGRPLGHTDAWGLAYSAHQRLISC 587

Search completed: July 24, 2003, 17:06:35
Job time : 26 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:03:37 ; Search time 94 Seconds

(without alignments)
466.691 Million cell updates/sec

Title: US-09-171-607A-1

Perfect score: 893

Sequence: 1 VALNSPLSGMGRGIRGADFO.....ASCHHAYIVLCIENSFTMAS 170

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	893	100.0	187	4	OBWXS15
2	893	100.0	261	4	OBWXS15
3	893	100.0	816	4	OBWXS15
4	778	87.1	184	11	OBWXS15
5	778	87.1	1140	11	OBWXS15
6	778	87.1	1174	11	OBWXS15
7	763	85.4	226	11	OBWXS15
8	757	84.8	171	11	OBWXS15
9	722	80.9	160	11	OBWXS15
10	711	79.6	1344	13	OBWXS15
11	654	73.2	1315	13	OBWXS15
12	643	72.0	1307	13	OBWXS15
13	606	67.9	361	13	OBWXS15
14	529	59.2	102	4	OBWXS15
15	515	57.7	1367	11	OBWXS15
16	512	57.3	1367	11	OBWXS15

17	505	56.6	1388	4	Q9Y4W4	Q9Y4W4 homo sapien
18	424.5	47.5	292	5	OBMS3	OBMS3 drosophila
19	418.5	46.9	792	5	OBMT89	OBMT89 drosophila
20	357	40.0	650	5	O17866	O17866 caenorhabd
21	357	40.0	778	5	O909K6	O909K6 caenorhabd
22	357	40.0	1117	5	O909K7	O909K7 caenorhabd
23	357	40.0	1154	5	OB10G6	OB10G6 caenorhabd
24	85	9.5	478	4	OB1V1	OB1V1 homo sapien
25	85	9.5	1649	4	OB1WY7	OB1WY7 homo sapien
26	82.5	9.2	208	16	Q9KX28	Q9KX28 rhizobium m
27	82	9.2	498	16	Q9KX2	Q9KX2 streptomyce
28	80	9.0	493	16	O98AT9	O98AT9 rhizobium 1
29	80	9.0	651	5	O9VFA9	O9VFA9 drosophila
30	80	9.0	1024	16	O8FRH1	O8FRH1 corynebacte
31	79.5	8.9	240	16	O8FR63	O8FR63 leptospira
32	79.5	8.9	716	10	OB1J38	OB1J38 oryza sativ
33	79	8.8	477	10	O9SMY7	O9SMY7 arabidopsis
34	79	8.8	525	10	O94JL8	O94JL8 arabidopsis
35	78.5	8.8	285	16	O98E05	O98E05 rhizobium 1
36	78	8.7	1715	6	O9GLM4	O9GLM4 bos taurus
37	77.5	8.7	395	11	O9QUP4	O9QUP4 mus musculu
38	77	8.6	314	16	O8U8N8	O8U8N8 agrobacteri
39	76	8.5	204	4	O96GA9	O96GA9 homo sapien
40	76	8.5	204	11	O8JZR2	O8JZR2 mus musculu
41	76	8.5	636	4	O60624	O60624 homo sapien
42	76	8.5	1702	12	OB1JX5	OB1JX5 normal-k-lik
43	76	8.5	1702	12	OB1JX4	OB1JX4 normal-k-lik
44	75.5	8.5	244	16	O8EAW6	O8EAW6 shewanella
45	75.5	8.5	334	11	O9WTJ8	O9WTJ8 mus musculu

ALIGNMENTS

RESULT 1
OBWXS15 PRELIMINARY; PRT: 187 AA.

AC OBWXS15
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE Collagen XVIII (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21409408; PubMed=11517600;
RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
RT "Inhibition effect in vitro of purified endostatin expressed in Pichia
pastoris".
RL Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
DR EMBL; AF416592; AAL37720.1; -;
FT NON-TER
SQ SEQUENCE 187 AA: 20448 MM: 72B1047D85838CD3 CRC64;

Query Match 100.0%; Score 893; DB 4; Length 187;
Best Local Similarity 100.0%; Pred. No. 4.9e-83;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRATLSSRLDLYSVRRADRAVPYV 60
DB 17 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRATLSSRLDLYSVRRADRAVPYV 76
QY 61 NIKDELFPSWEALFSSSEGLPKGARIFSPDGKDVLRHPWPQKSWHGSDPGRRLTE 120
DB 77 NIKDELFPSWEALFSSSEGLPKGARIFSPDGKDVLRHPWPQKSWHGSDPGRRLTE 136
QY 121 SYCETWTEAPSATGQASSLLGRLGQASASCHHAYIVLCIENSFTMAS 170
DB 137 SYCETWTEAPSATGQASSLLGRLGQASASCHHAYIVLCIENSFTMAS 186

RESULT 2

08NG19 PRELIMINARY: PRT: 261 AA.

BT 01-OCT-2002 (TRENBLREL. 22, Created)
DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
DE 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Multi-functional protein MFP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF282883; AAM52249.1; -
SO SEQUENCE 261 AA; 26745 MW; CA60C920AF3E90E5 CRC64;

Query Match 100.0%; Score 893; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 7, 4e-83;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRIGADFCFOQARAAGLAGTFRALSSRLQDLYSIVRADRAAPVIV 60
DB 91 VALNSPLSGMGRIGADFCFOQARAAGLAGTFRALSSRLQDLYSIVRADRAAPVIV 150
OY 61 NLKDELLFPSSWEALFSGSEGGLPKGARIFESFDGKDVLRHPWPQKSWHGSDDPNGRRLTE 120
DB 151 NLKDELLFPSSWEALFSGSEGGLPKGARIFESFDGKDVLRHPWPQKSWHGSDDPNGRRLTE 210
OY 121 SYCETWREAPSATGQASSLLGGRLLGOSASCHHAYIVLCIENSFMTAS 170
DB 211 SYCETWREAPSATGQASSLLGGRLLGOSASCHHAYIVLCIENSFMTAS 260

RESULT 3

08N4S4 PRELIMINARY: PRT: 816 AA.

BT 01-OCT-2002 (TRENBLREL. 22, Created)
DT 01-OCT-2002 (TRENBLREL. 22, Last sequence update)
DE 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Similar to collagen, type XVIII, alpha 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Renal adenocarcinoma;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC033715; AAB33715.1; -
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 6.
DR ProDom: PD000007; Collagen; 1.
KW Collagen.
FT NON_TER 1
SO SEQUENCE 816 AA; 82553 MW; 50539B2946694F86 CRC64;

Query Match 100.0%; Score 893; DB 4; Length 816;
Best Local Similarity 100.0%; Pred. No. 3, 1e-82;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRIGADFCFOQARAAGLAGTFRALSSRLQDLYSIVRADRAAPVIV 60
DB 646 VALNSPLSGMGRIGADFCFOQARAAGLAGTFRALSSRLQDLYSIVRADRAAPVIV 705
OY 61 NLKDELLFPSSWEALFSGSEGGLPKGARIFESFDGKDVLRHPWPQKSWHGSDDPNGRRLTE 120
DB 706 NLKDELLFPSSWEALFSGSEGGLPKGARIFESFDGKDVLRHPWPQKSWHGSDDPNGRRLTE 765

OY 121 SYCETWREAPSATGQASSLLGGRLLGOSASCHHAYIVLCIENSFMTAS 170
DB 766 SYCETWREAPSATGQASSLLGGRLLGOSASCHHAYIVLCIENSFMTAS 815

RESULT 4

09JK63 PRELIMINARY: PRT: 184 AA.

BT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DE 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE Endostatin (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Chinese Kunming;
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor neovasculature."
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF257775; AAF69009.1; -
DR HSSP: P39061; IKOE.
FT NON_TER 1
FT NON_TER 1
SO SEQUENCE 184 AA; 20376 MW; AC06F9D8D103412A CRC64;

Query Match 87.1%; Score 778; DB 11; Length 184;
Best Local Similarity 85.8%; Pred. No. 2, 5e-71;
Matches 145; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRIGADFCFOQARAAGLAGTFRALSSRLQDLYSIVRADRAAPVIV 60
DB 13 VALNSPLSGMGRIGADFCFOQARAAGLAGTFRALSSRLQDLYSIVRADRAAPVIV 72
OY 61 NLKDELLFPSSWEALFSGSEGGLPKGARIFESFDGKDVLRHPWPQKSWHGSDDPNGRRLTE 120
DB 73 NLKDELLFPSSWEALFSGSEGGLPKGARIFESFDGKDVLRHPWPQKSWHGSDDPNGRRLTE 132
OY 121 SYCETWREAPSATGQASSLLGGRLLGOSASCHHAYIVLCIENSFMTA 169
DB 133 SYCETWREAPSATGQASSLLGGRLLGOSASCHHAYIVLCIENSFMTAS 181

RESULT 5

061434 PRELIMINARY: PRT: 1140 AA.

BT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioke H., Inoue H., Niinomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196; 576-582 (1993).
DR EMBL: D17546; BAA04483.1; -
DR HSSP: P39061; IKOE.
DR MCD: MGI:88449; COL15A1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF01391; Collagen; 8.

QY 121 SYCETWTEAPSATGQASSLLGRLGQASASCHHAYIVLCIENSFMTA 169
 DB 175 SYCETWTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFMTS 223

RESULT 8

Q9WUM5 PRELIMINARY; PRT; 171 AA.

AC Q9WUM5;
 ID 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Collagen type XVIII, alpha (I) chain (Fragment).
 GN COL18A1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_Taxid=10116;

(1) SEQUENCE FROM N.A.

RC TISSUE-Liver;
 RA Jia J.D., Bauer M., Eberspacher U., Donner P., Schuppan D.;
 RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.

(2) SEQUENCE FROM N.A.

RC TISSUE-Liver;
 RA Jia J.D., Bauer M., Sedlaczek N., Ruehl M., Riecken E.O., Schuppan D.;
 RT "Temporopartial expression of collagen XVIII/endostatin in acute and
 chronic liver injuries.";
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ236873; CAB44263.1; -

DR HSSP; P39061; IKOE.

FT NON_TER 1 1

FT NON_TER 171 171

SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

Query Match 84.8%; Score 757; DB 11; Length 171;
 Best Local Similarity 85.0%; Pred. No. 3, 2e-69;
 Matches 142; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGRIGRAGDFQCFQOARAVAGLAGTFRAFLSSRLQDLYSIVRRADRAVPIV 60

DB 5 VALNTPLSGMGRIGRAGDFQCFQOARAVAGLSGTFRFLSSRLQDLYSIVRRADSSVPIV 64

QY 61 NLKDELFPSWEALFSGSEGLPKGARIFSFQDKDVLRHPTWPKSWHSGSDPNGRRLTE 120

DB 65 NLKDEVLPSPMDTLFSGSQGLHSGARIFSFQDRDVLRHPTWPKSWHSGSDPNGRRLME 124

QY 121 SYCETWTEAPSATGQASSLLGRLGQASASCHHAYIVLCIENSFMT 167

DB 125 SYCETWTEATGVTGQASSLLSGRLLEQKAESCHNSYIVLCIENSFMT 171

RESULT 9

Q9CRT2 PRELIMINARY; PRT; 160 AA.

AC Q9CRT2;
 ID 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Procollagen, type XVIII, alpha 1 (Fragment).
 GN COL18A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_Taxid=10090;

(1) SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.T., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Williams L.,
 RA Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690 (2001).
 DR EMBL; AK014292; BAB29249.1; -

DR HSSP; P39061; IKOE.
 DR MGD; MGI:88451; COL18A1.
 FT NON_TER 1 1
 SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;

Query Match 80.9%; Score 722; DB 11; Length 160;
 Best Local Similarity 85.4%; Pred. No. 1, 1e-65;
 Matches 134; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 13 GIRGADFCFOQARAVAGLAGTFRAFLSSRLQDLYSIVRRADRAVPIVNLDELFPSE 72

DB 1 GIRGADFCFOQARAVAGLSGTFRFLSSRLQDLYSIVRRADRGSPYIYNLDEVLPSPMD 60

QY 73 ALFSGSEGLPKGARIFSFQDKDVLRHPTWPKSWHSGSDPNGRRLTESYCETWTEAPS 132

DB 61 SLFSGSQQLPGARIFSFQDRDVLRHPTWPKSWHSGSDPNGRRLTESYCETWTEATTG 120

QY 133 ATGQASSLLSGRLGQASASCHHAYIVLCIENSFMTA 169

DB 121 ATGQASSLLSGRLLEQKAESCHNSYIVLCIENSFMTS 157

RESULT 10

Q93419 PRELIMINARY; PRT; 1344 AA.

AC Q93419;
 ID 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Collagen XVIII precursor.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_Taxid=9031;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Halfter W., Dong S., Schurer B., Cole G.J.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF083440; AAC33294.2; -

DR HSSP; P39061; IKOE.

DR InterPro; IPR0006087; Collagen.

DR InterPro; IPR001791; Laminin_G.

DR InterPro; IPR003129; TSPN.

DR Pfam; PF01391; Collagen; 8.

DR Pfam; PF02210; TSPN; 1.

DR ProDom; PD0000007; Collagen; 2.

DR SMART: SM00282; LamG: 1.
 DR SMART: SM00210; TSPN: 1.
 KW Collagen: Signal.
 FT SIGNAL
 SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;

Query Match 79.6%; Score 711; DB 13; Length 1344;
 Best Local Similarity 77.6%; Pred. No. 2.1e-63;
 Matches 132; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 60
 DB 1173 VALNPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 1232
 QY 61 NIKDELFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPWPQKSVHSGSPNGRRLTE 120
 DB 1233 NLRDEVLYDNWDSLFSGSEGLPKPGARIFSPDGKDVLRHPWPQKSVHSGSPNGRRLTE 1292
 QY 121 SYCETWRTAPSAATGOASSLLGRLGSAASCHHAYIVLCIENSFMTAS 170
 DB 1293 SYCEAWRTDERCTSGOASSLLGRLGSAASCHHAYIVLCIENSFMTAA 1342

RESULT 11

Q80HL9 PRELIMINARY; PRT; 1315 AA.
 AC Q80HL9;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Type XVIII collagen alpha chain.
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Xenopodinae; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 CC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ishino T., Sekimizu K., Natori S., Kubo T.;
 RT "Identification and characterization of genes expressed selectively in the regenerating tail of *Xenopus laevis* tadpole";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB047066; BAB84674.1;
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR003129; TSPN.
 DR Pfam: PF02210; TSPN; 1.
 DR ProDom: PD000007; Collagen; 1.
 DR SMART: SM00210; TSPN; 1.
 DR Collagen.
 SQ SEQUENCE 1315 AA; 134946 MW; 0C56C235DE058365 CRC64;

Query Match 73.2%; Score 654; DB 13; Length 1315;
 Best Local Similarity 73.2%; Pred. No. 1.3e-57;
 Matches 123; Conservative 16; Mismatches 29; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 60
 DB 1144 VALNPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 1203
 QY 61 NIKDELFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPWPQKSVHSGSPNGRRLTE 120
 DB 1204 NLRDEVLYDNWDSLFSGSEGLPKPGARIFSPDGKDVLRHPWPQKSVHSGSPNGRRLTE 1263
 QY 121 SYCETWRTAPSAATGOASSLLGRLGSAASCHHAYIVLCIENSFMTAS 168
 DB 1264 SYCETWRTAPSAATGOASSLLGRLGSAASCHHAYIVLCIENSFMTA 1311

RESULT 12
 ID Q8JFE7 PRELIMINARY; PRT; 1307 AA.
 AC Q8JFE7;
 ID Q8JFE7;
 AC Q8JFE7;

DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Type XVIII collagen short variant.
 OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 CC Xenopodinae; Xenopus.
 CC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-22166979; PubMed-12175494;
 RA Elama H., Peterson J., Pihlajaniemi T., Destree O.;
 RT "Cloning of three variants of type XVIII collagen and their expression patterns during *Xenopus laevis* development";
 RL Mech. Dev. 114:109-113(2002).
 DR EMBL: AY052763; AL14257.1;
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR003129; TSPN.
 DR Pfam: PF02210; TSPN; 1.
 DR SMART: SM00210; TSPN; 1.
 DR Collagen.
 SQ SEQUENCE 1307 AA; 134153 MW; D53EDBE3DE34976 CRC64;

Query Match 72.0%; Score 643; DB 13; Length 1307;
 Best Local Similarity 71.4%; Pred. No. 1.7e-56;
 Matches 120; Conservative 18; Mismatches 30; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 60
 DB 1136 VALNPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 1195
 QY 61 NIKDELFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPWPQKSVHSGSPNGRRLTE 120
 DB 1196 NLRDEVLYDNWDSLFSGSEGLPKPGARIFSPDGKDVLRHPWPQKSVHSGSPNGRRLTE 1255
 QY 121 SYCETWRTAPSAATGOASSLLGRLGSAASCHHAYIVLCIENSFMTAS 168
 DB 1256 SYCETWRTAPSAATGOASSLLGRLGSAASCHHAYIVLCIENSFMTA 1303

RESULT 13

Q8AWC6 PRELIMINARY; PRT; 361 AA.
 AC Q8AWC6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Collagen XVIII (Fragment).
 OS Branchydanio rerio (zebrafish) (Danio rerio).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 CC Cyprinidae; Danio.
 CC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haftik Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
 RT "Sequence and embryonic expression of collagen XVIII NC1 domain (endostatin) in the zebrafish";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ494837; CAD38825.1;
 DR NON_TER
 SQ SEQUENCE 361 AA; 40222 MW; 3C5A0F8479D26735 CRC64;

Query Match 67.9%; Score 606; DB 13; Length 361;
 Best Local Similarity 66.5%; Pred. No. 2e-53;
 Matches 113; Conservative 22; Mismatches 35; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 60
 DB 190 IALNSPQVGMIRGADFCQFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAVPY 249

OY 61 NLKDELFPSSWEALFSGSEGPLKPGARIFSEFDGKDVLRHPTWPKSVHSGSDPNGRRLTE 120
 DB 250 NLKQGVLFSSWESLFSDESRMKDNAPITYSFDGRVLDNSANPEMIMHSGSDGRHROT 309
 OY 121 SYCETWRTAPSPATGQASSLLGRLGSAASCHHAYIVLCIENSFMTAS 170
 DB 310 NYCETWRAGDRAVYGTGLASSLQAGLLQOTSSSSSSSYIALCIENSXYMTOS 359

RESULT 14

096T70 PRELIMINARY; PRT; 102 AA.
 AC 096T70;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Endostatin variant (Fragment).
 SS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Deininger M.H., Trautmann K., Schluesener H.J.;
 RT "Endostatin promotes delayed secondary damage following traumatic
 brain injury."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF333247; AAK50626.1; -.
 FT NON_TER 1
 FT 102 1
 SQ SEQUENCE 102 AA; 11147 MW; ECACA47AA6420947D CRC64;

Query Match 59.2%; Score 529; DB 4; Length 102;
 Best Local Similarity 98.0%; Pred. No. 2.8e-46;
 Matches 100; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 59 IVNLKDELFPSSWEALFSGSEGPLKPGARIFSEFDGKDVLRHPTWPKSVHSGSDPNGRRL 118
 DB 1 IVNLKDELFPSSWEALFSGSEGPLKPGARIFSEFDGKDVLRHPTWPKSVHSGSDPNGRRL 60
 OY 119 TESTCETWRTAPSPATGQASSLLGRLGSAASCHHAYIVL 160
 DB 61 TESTCETWRTAPSPATGQASSLLGRLGSAASCHHAYIVL 102

RESULT 15

035206 PRELIMINARY; PRT; 1367 AA.
 AC 035206;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Type XV collagen.
 GN COL15A1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97480713; PubMed=9339358;
 RA Haeg P.M., Horelli-Kutunnen N., Eklund L., Palotie A.,
 RA Pihlajaniemi T.;
 RT "Cloning of mouse type XV collagen sequences and mapping of the
 RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
 RT (XV) collagen sequences indicates divergence in the number of small
 RT collagenous domains."
 RL Genomics 45:31-41(1997).
 DR EMBL; AF011450; AAC53387.1; -.
 DR HSSP; P39061; IKOE.
 DR MGD; MGI:88449; COL15A1.
 DR InterPro; IPR000087; Collagen.

DR InterPro; IPR001791; Laminin_G.
 DR InterPro; IPR003129; TSPN.
 DR Pfam; PF01391; Collagen; 5.
 DR Pfam; PF02210; TSPN; 1.
 DR ProDom; PD000007; Collagen; 1.
 DR SMART; SM00282; LamG; 1.
 DR SMART; SM00210; TSPN; 1.
 KW Collagen.
 SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3AEEC CRC64;

Query Match 57.7%; Score 515; DB 11; Length 1367;
 Best Local Similarity 58.3%; Pred. No. 2e-43;
 Matches 98; Conservative 26; Mismatches 40; Indels 4; Gaps 1;

OY 1 VALNSPLSGMRGIRGADFOCFQARAYGLAGTERAFISSLRLQDLYSVRRADRAAPIV 60
 DB 1200 VALNPVAVGDIR---ADFOCFQARAYGLSTFRAFLSHLQDLYSVRRARERGDPIV 1255
 OY 61 NLKDELFPSSWEALFSGSEGPLKPGARIFSEFDGKDVLRHPTWPKSVHSGSDPNGRRLTE 120
 DB 1256 NLKQGVLFNNMDSIFSGDGGQFNTHIPIYSFDGRVMDPSWPKVYVHSGNPHGRLVD 1315
 OY 121 SYCETWRTAPSPATGQASSLLGRLGSAASCHHAYIVLCIENSFMT 168
 DB 1316 KYCEAWRTTDMAVYGFASPLSTGKILDKAYSCANRLIVLCIENSFMT 1363

Search completed: July 24, 2003, 17:08:18
 Job time : 99 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:08:22 ; Search time 52 Seconds
(without alignments)
388,253 Million cell updates/sec

Title: US-09-171-607a-1

Perfect score: 893

Sequence: 1 VALNSPLSGMGRGADPQ.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications, AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	893	100.0	178	US-10-131-241-60	Sequence 60, Appl
2	893	100.0	178	US-10-042-347-5	Sequence 5, Appl
3	893	100.0	179	US-10-131-241-57	Sequence 57, Appl
4	893	100.0	182	US-09-998-831-14	Sequence 14, Appl
5	893	100.0	182	US-10-131-241-54	Sequence 54, Appl
6	893	100.0	182	US-10-042-347-3	Sequence 3, Appl
7	893	100.0	183	US-09-873-676-2	Sequence 2, Appl
8	893	100.0	183	US-10-080-797-1	Sequence 1, Appl
9	893	100.0	183	US-10-131-241-52	Sequence 52, Appl
10	893	100.0	184	US-09-961-403-5	Sequence 5, Appl
11	893	100.0	1516	US-10-060-036-166	Sequence 166, App
12	889	99.6	181	US-10-131-241-55	Sequence 55, Appl
13	885	99.1	180	US-10-131-241-56	Sequence 56, Appl
14	858	96.1	180	US-10-131-241-47	Sequence 47, Appl
15	783	87.7	184	US-10-131-241-49	Sequence 49, Appl

16	778	87.1	191	US-09-998-831-13	Sequence 13, Appl
17	778	87.1	207	US-10-080-797-3	Sequence 3, Appl
18	775	86.8	184	US-10-131-241-46	Sequence 46, Appl
19	734	82.2	185	US-10-036-869-36	Sequence 36, Appl
20	346	38.7	63	US-09-822-540A-1	Sequence 1, Appl
21	182	20.4	31	US-09-822-540A-2	Sequence 2, Appl
22	94	10.5	16	US-09-766-412-32	Sequence 32, Appl
23	85	9.5	10	US-09-815-915-8	Sequence 8, Appl
24	80	9.0	409	US-10-156-761-13624	Sequence 13624, A
25	77.5	8.7	418	US-09-927-602-5	Sequence 5, Appl
26	76.5	8.6	421	US-10-156-761-13612	Sequence 13612, A
27	75	8.4	332	US-09-815-242-5659	Sequence 5659, Ap
28	75	8.4	346	US-09-815-242-12271	Sequence 12271, A
29	73.5	8.2	1653	US-09-741-669-402	Sequence 402, App
30	73	8.2	582	US-10-156-761-11262	Sequence 11262, A
31	73	8.2	690	US-10-102-806-726	Sequence 726, App
32	72	8.1	304	US-09-879-957-28	Sequence 28, Appl
33	72	8.1	636	US-10-000-776-12	Sequence 12, Appl
34	72	8.1	636	US-10-156-761-11553	Sequence 11553, A
35	72	8.1	745	US-09-866-050A-148	Sequence 148, App
36	70	7.8	210	US-10-156-761-9287	Sequence 9287, Ap
37	70	7.8	229	US-10-156-761-13097	Sequence 13097, A
38	70	7.8	739	US-09-728-721-43	Sequence 43, Appl
39	70	7.8	953	US-10-118-984-43	Sequence 43, Appl
40	70	7.8	953	US-10-295-981-43	Sequence 43, Appl
41	70	7.8	1506	US-10-128-714-3216	Sequence 3216, Ap
42	69.5	7.8	1506	US-10-128-714-8216	Sequence 8216, Ap
43	69.5	7.8	1313	US-09-883-691-6	Sequence 6, Appl
44	69	7.7	318	US-10-146-772-118	Sequence 118, App
45	69	7.7	15		

ALIGNMENTS

RESULT 1
US-10-131-241-60
Sequence 60, Application US/10131241
Publication No. US20030012792A1
GENERAL INFORMATION:
APPLICANT: Holaday, John W.
APPLICANT: Fortier, Anne H.
TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prol
FILE REFERENCE: 05213-0344 43170-271565
CURRENT APPLICATION NUMBER: US/10/131,241
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US 09/413,049
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: US 09/316,802
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: US 60/086,586
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 60
LENGTH: 178
TYPE: PRT
ORGANISM: Homo sapiens
US-10-131-241-60
Query Match 100.0% Score 893; DB 15; Length 178;
Best Local Similarity 100.0% Pred. No. 4.8e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 VALNSPLSGMGRGADPQFOGQARAVGLAGTFRATLSSRLQDLYSTRADAAVIV 60
DB 9 VALNSPLSGMGRGADPQFOGQARAVGLAGTFRATLSSRLQDLYSTRADAAVIV 68
OY 61 NLKDELLFPPSEALFSSSEGLPKGARIFFSDGADVLRHPMPQKSVYHGSDPNRRLTE 120
DB 69 NLKDELLFPPSEALFSSSEGLPKGARIFFSDGADVLRHPMPQKSVYHGSDPNRRLTE 128

QY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
DB 129 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 178

RESULT 2

US-10-042-347-5
Sequence 5, Application US/10042347
Publication No. US20030114370A1

GENERAL INFORMATION:

APPLICANT: O'Reilly, Michael S.

APPLICANT: Folkman, M. Judah

TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fr

TITLE OF INVENTION: Thereof

FILE REFERENCE: 05213-0880 (43170-249874)

CURRENT APPLICATION NUMBER: US/10/042,347

CURRENT FILING DATE: 2002-01-11

PRIOR APPLICATION NUMBER: US 09/315,689

PRIOR FILING DATE: 1999-05-20

PRIOR APPLICATION NUMBER: US 60/106,343

PRIOR FILING DATE: 1998-10-30

PRIOR APPLICATION NUMBER: US 09/154,302

PRIOR FILING DATE: 1998-09-16

PRIOR APPLICATION NUMBER: US 08/740,168

PRIOR FILING DATE: 1996-10-22

PRIOR APPLICATION NUMBER: US 60/005,835

PRIOR FILING DATE: 1995-10-23

PRIOR APPLICATION NUMBER: US 60/023,070

PRIOR FILING DATE: 1996-08-02

PRIOR APPLICATION NUMBER: US 60/026,263

PRIOR FILING DATE: 1996-09-17

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 178

TYPE: PRT

ORGANISM: Homo sapiens

US-10-042-347-5

Query Match Best Local Similarity 100.0%; Score 893; DB 15; Length 178;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQCARAVGLAGTFRAFLLSSRLQDLYSIVRRADRAAPV 60

DB 9 VALNSPLSGMGIRGADFOCFQCARAVGLAGTFRAFLLSSRLQDLYSIVRRADRAAPV 68

61 NLKDELLPSPWEALFSSGSEGPLKPGARIFSPDGKIVLHPTWPKSVWHGSDPNRRRLTE 120

DB 69 NLKDELLPSPWEALFSSGSEGPLKPGARIFSPDGKIVLHPTWPKSVWHGSDPNRRRLTE 128

QY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170

DB 129 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 178

RESULT 3

US-10-131-241-57

Sequence 57, Application US/10131241

Publication No. US20030012792A1

GENERAL INFORMATION:

APPLICANT: Holaday, John W.

APPLICANT: Fortier, Anne H.

TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer

FILE REFERENCE: 05213-0344 43170-271565

CURRENT APPLICATION NUMBER: US/10/131,241

CURRENT FILING DATE: 2002-07-22

PRIOR APPLICATION NUMBER: US 09/413,049

PRIOR FILING DATE: 1999-10-06

PRIOR APPLICATION NUMBER: US 09/316,802

PRIOR FILING DATE: 1999-05-21

PRIOR APPLICATION NUMBER: US 60/086,586

PRIOR FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO 57
LENGTH: 179
TYPE: PRT
ORGANISM: Homo sapiens
US-10-131-241-57

Query Match

Best Local Similarity 100.0%; Score 893; DB 15; Length 179;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQCARAVGLAGTFRAFLLSSRLQDLYSIVRRADRAAPV 60

DB 9 VALNSPLSGMGIRGADFOCFQCARAVGLAGTFRAFLLSSRLQDLYSIVRRADRAAPV 68

61 NLKDELLPSPWEALFSSGSEGPLKPGARIFSPDGKIVLHPTWPKSVWHGSDPNRRRLTE 120

DB 69 NLKDELLPSPWEALFSSGSEGPLKPGARIFSPDGKIVLHPTWPKSVWHGSDPNRRRLTE 128

QY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170

DB 129 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 178

RESULT 4

US-09-998-831-14

Sequence 14, Application US/09998831

Patent No. US20020119153A1

GENERAL INFORMATION:

APPLICANT: Phillip E. Thorpe

APPLICANT: Rolf A. Brecken

TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY

FILE REFERENCE: 4001.002584

CURRENT APPLICATION NUMBER: US/09/998,831

CURRENT FILING DATE: 2001-11-30

PRIOR APPLICATION NUMBER: 09/561,108

PRIOR FILING DATE: 2000-04-28

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 182

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-998-831-14

Query Match Best Local Similarity 100.0%; Score 893; DB 10; Length 182;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQCARAVGLAGTFRAFLLSSRLQDLYSIVRRADRAAPV 60

DB 13 VALNSPLSGMGIRGADFOCFQCARAVGLAGTFRAFLLSSRLQDLYSIVRRADRAAPV 72

QY 61 NLKDELLPSPWEALFSSGSEGPLKPGARIFSPDGKIVLHPTWPKSVWHGSDPNRRRLTE 120

DB 73 NLKDELLPSPWEALFSSGSEGPLKPGARIFSPDGKIVLHPTWPKSVWHGSDPNRRRLTE 132

QY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170

DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 5

US-10-131-241-54

Sequence 54, Application US/10131241

Publication No. US20030012792A1

GENERAL INFORMATION:

APPLICANT: Holaday, John W.
APPLICANT: Fortler, Anne H.
TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Proliferation
FILE REFERENCE: 05213-0344 43170-271565
CURRENT APPLICATION NUMBER: US/10/131,241
PRIOR FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US 09/413,049
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: US 09/316,802
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: US 60/086,586
PRIOR FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patentin version 3.1
SEQ ID NO 54
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-10-131-241-54

Query Match 100.0%; Score 893; DB 15; Length 182;
Best Local Similarity 100.0%; Pred. No. 4.9e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLACTFRFAFLSSRLQDLYSIYRRADRAAVPIY 60
DB 13 VALNSPLSGMGIRGADFOCFQOARAVGLACTFRFAFLSSRLQDLYSIYRRADRAAVPIY 72
QY 61 NIKDELFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPWPQKSVHSGSDPNGRRLTE 120
DB 73 NIKDELFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPWPQKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 6
US-10-042-347-3
Sequence 3, Application US/10042347
Publication No. US20030114370A1
GENERAL INFORMATION:
APPLICANT: O'Reilly, Michael S.
APPLICANT: Folkman, M. Judah
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fragments
FILE REFERENCE: 05213-0880 (43170-249874)
CURRENT APPLICATION NUMBER: US/10/042,347
PRIOR FILING DATE: 2002-01-11
PRIOR APPLICATION NUMBER: US 09/315,689
PRIOR FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: US 60/106,343
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: US 09/154,302
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: US 08/740,168
PRIOR FILING DATE: 1996-10-22
PRIOR APPLICATION NUMBER: US 60/005,835
PRIOR FILING DATE: 1995-10-23
PRIOR APPLICATION NUMBER: US 60/023,070
PRIOR FILING DATE: 1996-08-02
PRIOR APPLICATION NUMBER: US 60/026,263
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin version 3.1
SEQ ID NO 3
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-10-042-347-3

Query Match 100.0%; Score 893; DB 15; Length 182;

Best Local Similarity 100.0%; Pred. No. 4.9e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLACTFRFAFLSSRLQDLYSIYRRADRAAVPIY 60
DB 13 VALNSPLSGMGIRGADFOCFQOARAVGLACTFRFAFLSSRLQDLYSIYRRADRAAVPIY 72
QY 61 NIKDELFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPWPQKSVHSGSDPNGRRLTE 120
DB 73 NIKDELFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPWPQKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 7
US-09-873-676-2
Sequence 2, Application US/09873676
Patent No. US20020077289A1
GENERAL INFORMATION:
APPLICANT: Macdonald, Nicholas J.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use
FILE REFERENCE: 05213-0378 (43170-259333)
CURRENT APPLICATION NUMBER: US/09/873,676
PRIOR FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: US 60/209,065
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: US 60/289,387
PRIOR FILING DATE: 2001-05-08
NUMBER OF SEQ ID NOS: 123
SOFTWARE: Patentin version 3.1
SEQ ID NO 2
LENGTH: 183
TYPE: PRT
ORGANISM: Homo sapiens
US-09-873-676-2

Query Match 100.0%; Score 893; DB 9; Length 183;
Best Local Similarity 100.0%; Pred. No. 4.9e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLACTFRFAFLSSRLQDLYSIYRRADRAAVPIY 60
DB 13 VALNSPLSGMGIRGADFOCFQOARAVGLACTFRFAFLSSRLQDLYSIYRRADRAAVPIY 72
QY 61 NIKDELFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPWPQKSVHSGSDPNGRRLTE 120
DB 73 NIKDELFPSSWEALFSGSGEPLKPGARIFSPDGKDVLRHPWPQKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 8
US-10-080-797-1
Sequence 1, Application US/10080797
Publication No. US20020183253A1
GENERAL INFORMATION:
APPLICANT: Campochiaro, Peter A.
APPLICANT: Dixon, Katharine H.
APPLICANT: Brazzell, Romulus K.
TITLE OF INVENTION: METHOD FOR TREATING OCULAR
FILE REFERENCE: 4-31881A
CURRENT APPLICATION NUMBER: US/10/080,797
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 183

TYPE: PRT
ORGANISM: Human
US-10-080-797-1

Query Match 100.0%; Score 893; DB 14; Length 183;
Best Local Similarity 100.0%; Pred. No. 4.9e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 60
DB 13 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 72
QY 61 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 73 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 182

RESULT 9

US-10-131-241-52

Sequence 52, Application US/10131241
Publication No. US20030012792A1
GENERAL INFORMATION:
APPLICANT: Holaday, John W.
TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Proliferation
FILE REFERENCE: 05213-0344 43170-271565
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: US 09/413,049
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: US 09/316,802
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: US 60/086,586
PRIOR FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO 52
LENGTH: 183
TYPE: PRT
ORGANISM: Homo sapiens
US-10-131-241-52

Query Match 100.0%; Score 893; DB 15; Length 183;
Best Local Similarity 100.0%; Pred. No. 4.9e-92;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 60
DB 13 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 72
QY 61 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 73 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 182

RESULT 10

US-09-961-403-5

Sequence 5, Application US/09961403
Publication No. US20030077589A1
GENERAL INFORMATION:
APPLICANT: HE-STUMP, HOLGER
APPLICANT: HAENDLER, BERNARD
APPLICANT: KRAETZSCHMAR, JOERN
APPLICANT: KREFT, BERTHOLD

APPLICANT: WINTERAGER, ELKE
APPLICANT: REGIDOR, PEDRO
APPLICANT: SCOTTI, SIMONE
TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
FILE REFERENCE: SCH-1789
CURRENT APPLICATION NUMBER: US/09/961,403
CURRENT FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 684
TYPE: PRT
ORGANISM: Homo sapiens
US-09-961-403-5

Query Match 100.0%; Score 893; DB 11; Length 684;
Best Local Similarity 100.0%; Pred. No. 2.7e-91;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 60
DB 514 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 573
QY 61 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 574 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 633
QY 121 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 170
DB 634 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 683

RESULT 11

US-10-060-036-166

Sequence 166, Application US/10060036
Publication No. US20030073144A1
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Kalos, Michael D.
APPLICANT: Lodes, Michael J.
APPLICANT: Persing, David H.
APPLICANT: Hepler, William T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.566
CURRENT APPLICATION NUMBER: US/10/060,036
CURRENT FILING DATE: 2002-01-30
NUMBER OF SEQ ID NOS: 4560
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 166
LENGTH: 1516
TYPE: PRT
ORGANISM: Homo sapiens
US-10-060-036-166

Query Match 100.0%; Score 893; DB 15; Length 1516;
Best Local Similarity 100.0%; Pred. No. 7.6e-91;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 60
DB 1346 VALNSPLSGMGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSIYRRADRAAPV 1405
QY 61 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 1406 NIKDELLFPPSWALEFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVHSGSDPNGRRLTE 1465
QY 121 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 170
DB 1466 SYCETWRTAPSATGQASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 1515

```
RESULT 12
US-10-131-241-55
; Sequence 55, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-55
```

```
Query Match          99.6%; Score 889; DB 15; Length 181;
Best Local Similarity 100.0%; Pred. No. 1,4e-91;
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLSSRLDLYSIVRRADRAAVPIV 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLSSRLDLYSIVRRADRAAVPIV 72
QY 61 NLKDELLFPPSWAEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 73 NLKDELLFPPSWAEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTSPATGQASSLLGRLGQSAASCHHAYIVLCIENSFMT 169
DB 133 SYCETWRTSPATGQASSLLGRLGQSAASCHHAYIVLCIENSFMT 181
```

```
RESULT 13
US-10-131-241-56
; Sequence 56, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-56
```

```
Query Match          99.1%; Score 885; DB 15; Length 180;
Best Local Similarity 100.0%; Pred. No. 3,8e-91;
Matches 168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLSSRLDLYSIVRRADRAAVPIV 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLSSRLDLYSIVRRADRAAVPIV 72
QY 61 NLKDELLFPPSWAEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 73 NLKDELLFPPSWAEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTSPATGQASSLLGRLGQSAASCHHAYIVLCIENSFMT 168
DB 133 SYCETWRTSPATGQASSLLGRLGQSAASCHHAYIVLCIENSFMT 180
```

```
RESULT 14
US-10-131-241-47
; Sequence 47, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prol
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Rhesus monkey
US-10-131-241-47
```

```
Query Match          96.1%; Score 858; DB 15; Length 180;
Best Local Similarity 95.8%; Pred. No. 4e-88;
Matches 161; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLSSRLDLYSIVRRADRAAVPIV 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLSSRLDLYSIVRRADRAAVPIV 72
QY 61 NLKDELLFPPSWAEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVHSGSDPNGRRLTE 120
DB 73 NLKDELLFPPSWAEALFSGSEGPLKPGARIFSPDKDVLRHPTWPKSVHSGSDPNGRRLTE 132
QY 121 SYCETWRTSPATGQASSLLGRLGQSAASCHHAYIVLCIENSFMT 168
DB 133 SYCETWRTSPATGQASSLLGRLGQSAASCHHAYIVLCIENSFMT 180
```

```
RESULT 15
US-10-131-241-49
; Sequence 49, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prol
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
```

; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 49
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Canine sp.
US-10-131-241-49

Query Match 87.7%; Score 783; DB 15; Length 184;
Best Local Similarity 85.8%; Pred No. 1e-79;
Matches 145; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

OY	1	VALNSPLSGMRGIRGADFCFOQARAYGLACTFRAFLSSRLQDLYSIVRRADRAVPY	60
Db	13	VALNSPQPGMRGIRGADFCFOQARAAGLACTFRAFLSSRLQDLYSIVRRADRTGVYV	72
OY	61	NKDELFPSSWALFSGSEGLKFGARIFSFQGDVLRHPTWPKSVWHSDDPNRRLTE	120
	73	NLRDEVLPSSWALFSGSEGLKFGARIFSFQGDVLRHPTWPKSVWHSDDPNRRLTE	132
OY	121	SYCETWRTPEAPATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMTA	169
Db	133	SYCETWRTPEAPATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMTA	181

Search completed: July 24, 2003, 17:17:17
Job time : 53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:04:02 ; Search time 39 Seconds

(without alignments)
419,197 Million cell updates/sec

Title: US-09-171-607A-1

Sequence: 1 VALNSPLSGMGKIRGADFO.....ASCHHAYIVLCIENSEFWTAS 170

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	893	100.0	684	2 A53019	collagen alpha 1(X
2	778	87.1	1774	2 B56101	collagen alpha 1(X
3	775	86.8	1335	2 A56101	collagen alpha 1(X
4	504	56.4	1388	2 A53317	collagen alpha 1(X
5	357	40.0	650	2 T22002	hypothetical prote
6	79	8.6	477	2 T05202	pectinesterase hom
7	77	8.6	314	2 AG3054	succinoglycan bios
8	77	8.6	314	2 F98231	succinoglycan bios
9	76	8.5	204	2 B45022	CRK-I - human
10	76	8.5	636	2 JMW047	class I cytochrome
11	75.5	8.5	904	2 T04377	probable pullulan
12	75	8.4	351	2 B89781	conserved hypothet
13	74.5	8.3	850	2 S56015	gastric mucin MUC5
14	73.5	8.2	1653	2 B91052	hypothetical prote
15	73.5	8.2	1653	2 F85896	hypothetical prote
16	73.5	8.2	1653	2 G65028	hypothetical prote
17	73	8.2	435	2 JC5584	hyaluronoglucosam
18	73	8.2	462	2 T17480	endo-xyranase homo
19	73	8.2	513	1 A45333	exopolysphatase
20	73	8.2	513	2 D91049	exopolysphatase
21	73	8.2	513	2 H85893	exopolysphatase
22	72.5	8.1	884	2 C70729	hypothetical prote
23	72	8.1	239	2 A46243	epidermal growth f
24	72	8.1	297	2 AE3128	carbohydrate kinas
25	72	8.1	304	2 A45022	CRK-II - human
26	72	8.1	304	2 I58394	c-Crk - mouse
27	72	8.1	319	2 D98159	hypothetical prote
28	72	8.1	800	2 AG0438	probable Rns acces
29	72	8.1	1006	2 T00050	hypothetical prote

30	71.5	8.0	554	2 T06374	probable pectinest
31	71.5	8.0	554	2 T06468	pectinesterase (EC
32	71	8.0	579	2 A70954	hypothetical prote
33	71	8.0	644	2 A72519	probable 2-oxoacid
34	71	8.0	732	2 AF0439	probable Rns acces
35	71	8.0	787	2 G81692	inner membrane pro
36	70.5	7.9	774	2 T04892	conserved hypothet
37	70	7.8	234	2 C75368	unsaturated glucur
38	70	7.8	375	2 F83906	hypothetical prote
39	70	7.8	451	2 H75593	hypothetical prote
40	70	7.8	672	2 T36083	conserved hypothet
41	70	7.8	741	2 AC0094	hypothetical prote
42	70	7.8	777	2 AD0982	conserved hypothet
43	69.5	7.8	168	2 S72898	biotin sulfoxide r
44	69.5	7.8	364	2 H87212	hypothetical prote
45	69	7.7	266	2 S22511	probable glucose e
					chlorophyll a/b-bi

ALIGNMENTS

RESULT 1
A53019
collagen alpha 1(XVIII) chain - human (fragment)
N:Contains: endostatin
C:Species: Homo sapiens (man)
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Mar-2000
C:Accession: A53019
R:Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse
Genomics 19, 494-499, 1994
A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and local
A:Reference number: A53019; MUID:94245237; PMID:8188291
A:Accession: A53019
A:Molecule type: mRNA
A:Residues: 1-684 <OHX>
A:Cross-references: GB:L22548; NID:9348908; PID:AA51864.1; PID:9562794
A>Note: The cited accession number, L25548, is not in Genbank release 103
A>Note: In the authors' translation, 482-Gly is not shown, residues 483-490 are shift
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
lated and subsequently O-glycosylated.
C:Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in periv
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of
ay be useful in treating solid tumors.
C:Genetics:
A:Gene: GDB:COL18A1
A:Cross-references: GDB:138752; OMIM:120328
A:Map position: 21q22.3-21q22.3
C:Superfamily: unassigned collagens
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
F:1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <KAT>
F:1-59/Domin: collagenous (fragment) #status predicted <CO4>
F:74-115/Domin: collagenous #status predicted <CO3>
F:129-201/Domin: collagenous #status predicted <CO6>
F:212-244/Domin: collagenous #status predicted <CO6>
F:257-278/Domin: collagenous #status predicted <CO7>
F:262-264/Region: cell attachment (R-G-D) motif
F:286-340/Region: collagenous #status predicted <CO8>
F:354-371/Domin: collagenous #status predicted <CO9>
F:502-684/Product: endostatin #status predicted <CO10>
F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 100.0%; Score 893; DB 2; Length 684;
Best local Similarity 100.0%; Pred. No. 1e-81; 0; Indels 0; Gaps 0;
Matches 170; Conservative 0; Mismatches 0

DB 514 VALNSPLSGMGKIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLTSIYRRADRAVPY 60
1 VALNSPLSGMGKIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLTSIYRRADRAVPY 60
|||
DB 514 VALNSPLSGMGKIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLTSIYRRADRAVPY 573
61 NKDELFPSPWALPFGSGRGLPKPARIRTSFGKDYVLRHPYTPQKSVHMGSDPNGRRLTE 120
|||
DB 574 NKDELFPSPWALPFGSGRGLPKPARIRTSFGKDYVLRHPYTPQKSVHMGSDPNGRRLTE 633

QY 121 SYCEWTRTEAPATGQASSLLGRLGGSASCHHAYIVLCIENSPMTAS 170
 DB 634 SYCEWTRTEAPATGQASSLLGRLGGSASCHHAYIVLCIENSPMTAS 683

RESULT 2

B56101
 collagen alpha 1(XVIII) chain precursor, long splice form - mouse
 N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
 C:Accession: B56101; S72450; S65595; PNO675; A54072; A58816
 R:Rehm, M.; PhilaJanleml, T.
 J. Biol. Chem. 270, 4705-4711, 1995
 A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue
 A:Reference number: A56101; MUID:95181468; PMID:7876242

A:Accession: B56101
 A:Molecule type: mRNA
 A:Residues: 1-562 <REH1>
 A:Cross-references: GB:U11637; NID:9618429; PIDN:AA52179.1; PID:9618430
 A:Experimental source: splice form clone PE17.24
 A:Accession: C56101

A:Molecule type: mRNA
 A:Residues: 1-239,487-562 <REH2>
 A:Cross-references: GB:U11637; NID:9618429
 A:Experimental source: splice form clones PE8.1, PE19, PE15.2
 R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
 submitted to the EMBL Data Library, August 1993
 A:Reference number: S72450

A:Accession: S72450
 A:Molecule type: mRNA
 A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16
 A:Cross-references: EMBL:L22345; NID:9448968; PIDN:AA19787.1; PID:9511298
 R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
 A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
 A:Reference number: A58370; MUID:94240111; PMID:8183893

A:Accession: S65595
 A:Molecule type: mRNA
 A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
 A:Cross-references: EMBL:L22345
 R:Abé, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Nishimura, Y.
 Biochem. Biophys. Res. Commun. 196, 576-582, 1993
 A:Title: Identification of a novel collagen chain represented by extensive interruptions
 A:Reference number: PNO675; MUID:94059075; PMID:8240330

A:Accession: PNO675
 A:Molecule type: mRNA
 A:Residues: 635-1774 <ABE>
 R:Rehm, M.; Hincikka, E.; PhilaJanleml, T.
 J. Biol. Chem. 269, 13929-13935, 1994
 A:Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str
 collagen chain.
 A:Reference number: A54072; MUID:9424507; PMID:8188673

A:Accession: A54072
 A:Molecule type: DNA; mRNA
 A:Residues: 1293-1403, 'R', 1405-1774 <REH3>
 A:Cross-references: GB:U03714; NID:9487733; PIDN:AA20657.1; PID:9487734
 R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B
 Cell 88, 277-285, 1997
 A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
 A:Reference number: A58816; MUID:97160848; PMID:9008168

A:Accession: A58816
 A:Molecule type: Protein
 A:Residues: 1591-1610 <REF>
 A:Experimental source: hemangioendothelium cells
 A:Note: Inhibits endothelial cell proliferation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit c
 lated and subsequently O-glycosylated.
 C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per
 C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
 A:Genetics: The useful in treating solid tumors.

A:Gene: MGI:Col18a1
 A:Cross-references: MGI:71175
 A:Map position: 10:41.0
 A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15
 A:Note: the list of introns is incomplete
 C:Superfamily: unassigned collagens
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog
 F:1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status p
 F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:361-486/Region: frizzled similarity
 F:786-812/Domain: collagenous #status predicted <CO01>
 F:823-896/Domain: collagenous #status predicted <CO02>
 F:921-1042/Domain: collagenous #status predicted <CO03>
 F:1066-1148/Domain: collagenous #status predicted <CO04>
 F:1163-1204/Domain: collagenous #status predicted <CO05>
 F:1218-1290/Domain: collagenous #status predicted <CO06>
 F:1301-1333/Domain: collagenous #status predicted <CO07>
 F:1346-1369/Domain: collagenous #status predicted <CO08>
 F:1351-1353/Region: cell attachment (R-G-D) motif
 F:1377-1428/Domain: collagenous #status predicted <CO09>
 F:1442-1459/Domain: collagenous #status predicted <CO10>
 F:1591-1774/Product: endostatin #status predicted <EST>
 F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
 F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 87.1%; Score 778; DB 2; Length 1774;
 Best Local Similarity 85.8%; Pred. No. 1,2e-69;
 Matches 145; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

QY 1 VALNSPLSGMKGIRGADFOCFQCARAVGLAGTFRAFLSSRLQDLYSTRADRAAVIV 60
 DB 1603 VALNPLSGMKGIRGADFOCFQCARAVGLAGTFRAFLSSRLQDLYSTRADRAAVIV 1662

QY 61 NIKDELPLPSSEALFSGSEGLPKPGARLFSPDGKVLNHPMPQKSVHSGSDPNKRRLTE 120
 DB 1663 NIKDELPLPSSEALFSGSEGLPKPGARLFSPDGKVLNHPMPQKSVHSGSDPNKRRLTE 1722

QY 121 SYCEWTRTEAPATGQASSLLGRLGGSASCHHAYIVLCIENSPMTA 169
 DB 1723 SYCEWTRTEAPATGQASSLLGRLGGSASCHHAYIVLCIENSPMTA 1771

RESULT 3

B56101
 collagen alpha 1(XVIII) chain precursor, short splice form - mouse
 N:Contains: endostatin
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
 C:Accession: A56101; A58371; S72450; S65595
 R:Rehm, M.; PhilaJanleml, T.
 J. Biol. Chem. 270, 4705-4711, 1995
 A:Title: Identification of three N-terminal ends of type XVIII collagen chains and ti
 A:Reference number: A56101; MUID:95181468; PMID:7876242

A:Accession: A56101
 A:Molecule type: mRNA
 A:Residues: 1-103 <REH1>
 A:Cross-references: GB:U11636; NID:9618427; PIDN:AA52178.1; PID:9618428
 R:Rehm, M.; PhilaJanleml, T.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
 A:Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collageno
 A:Reference number: A58371; MUID:94240112; PMID:8183894

A:Accession: A58371
 A:Molecule type: mRNA
 A:Residues: 1-928 <REH2>
 A:Cross-references: GB:U16898; NID:9404754; PIDN:AA37434.1; PID:9553894
 R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
 submitted to the EMBL Data Library, August 1993
 A:Reference number: S72450

A:Accession: S72450

A:Molecule type: mRNA
 A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>
 A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AAA19787.1; PID:9511298
 R:Och, S.P.; Kamagata, Y.; Murgaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
 A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
 A:Reference number: A58370; MUID:94240111; PMID:8183893
 A:Accession: S65595
 A:Molecule type: mRNA
 A:Residues: 28-1315 <OHS>
 A:Cross-references: EMBL:L22545
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 lated and subsequently O-glycosylated.
 C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per-
 C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un-
 C:Genetics:
 A:Gene: MGI:COL18A1
 A:Cross-references: MGI:71175
 A:Map position: 10:41.0
 A:Superfamily: unassigned collagens
 Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc-
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:24-235/Region: thrombospondin amino-terminal similarity
 F:26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <N
 F:327-353/Domain: collagenous #status predicted <CO1>
 F:364-437/Domain: collagenous #status predicted <CO2>
 F:462-583/Domain: collagenous #status predicted <CO3>
 F:607-689/Domain: collagenous #status predicted <CO4>
 F:704-745/Domain: collagenous #status predicted <CO5>
 F:759-831/Domain: collagenous #status predicted <CO6>
 F:847-874/Domain: collagenous #status predicted <CO7>
 F:887-910/Domain: collagenous #status predicted <CO8>
 F:892-894/Region: cell attachment (R-G-D) motif
 F:919-969/Domain: collagenous #status predicted <CO9>
 F:983-1000/Domain: collagenous #status predicted <CO10>
 F:1132-1315/Product: endostatin #status predicted <BST>
 F:1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
 F:120,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:177-228/Dissulfide bonds: #status predicted
 F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.8%; Score 775; DB 2; Length 1315;
 Best Local Similarity 85.2%; Pred. No. 1.7e-69;
 Matches 144; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRIGRADFOCFQOARAVGLAGTFRATLSRLQDLYSIVRRADRAVPYV 60
 |||||
 1144 VALNTPLSGGMRIGRADFOCFQOARAVGLSGTFRATLSRLQDLYSIVRRADRAVPYV 1203
 |||||

OY 61 NLADLFFPSWEALFSGSEGLPGARIFSPGKDVLRHPMPKSWHSGSDPGRRLTE 120
 |||||
 DB 1204 NLKDEVLSPSWDSLFSGSQGVQVGARIFSPGKDVLRHPMPKSWHSGSDPGRRLTE 1263
 |||||

OY 121 SYCETWTEAPSATGQASSLLGRLGQASASCHHAYIVLCIENSFMTA 169
 |||||
 DB 1264 SYCETWTEETTGATGQASSLLGRLGQASASCHHAYIVLCIENSFMTS 1312
 |||||

RESULT 4
 A53317
 collagen alpha 1(XV) chain precursor - human
 N:Alternate names: procollagen alpha 1(XV) chain
 C:Species: Homo sapiens (man)
 C:Date: 07-Jul-1995 #sequence, revision 07-Jul-1995 #text_change 31-Mar-2000
 C:Accession: A53317; A53146; S28778
 R:Kiviatko, S.; Heinemann, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.
 J. Biol. Chem. 269, 4773-4779, 1994
 A:Title: Primary structure of the alpha chain of human type XV collagen and exon-intron
 A:Reference number: A53317; MUID:9418920; PMID:8106446
 A:Accession: A53317
 A:Status: preliminary

A:Molecule type: mRNA
 A:Residues: 1-1388 <KIV>
 A:Cross-references: GB:I25280
 A:Note: nucleotide sequence and conceptual translation not complete
 R:Murgaki, Y.; Abe, N.; Minomiyu, Y.; Olsen, B.R.; Ooshima, A.
 J. Biol. Chem. 269, 4042-4046, 1994
 A:Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-trip
 A:Reference number: A53146; MUID:94140817; PMID:8307960
 A:Accession: A53146
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 1-9, 'S', 11-48, 'V', 50-94, 'A', 96-149, 'A', 151-203, 'V', 205-408, 'A', 410-569 <M
 A:Cross-references: GB:D21230; NID:9415605; PIDN:BA04762.1; PID:0105294; PID:946070
 R:Myers, J.C.; Kiviatko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.
 Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992
 A:Title: Identification of a previously unknown human collagen chain, alpha1(XV), cha
 A:Reference number: S28778; MUID:93066196; PMID:1279671
 A:Accession: S28778
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 544-640, 'P', 642-811, 'P', 813-1252 <MYE>
 A:Genetics:
 A:Gene: GDB:COL15A1
 A:Cross-references: GDB:132578; OMIM:120325
 A:Map position: 9q21-q22
 C:Superfamily: unassigned collagens
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>
 F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 56.4%; Score 504; DB 2; Length 1388;
 Best Local Similarity 56.9%; Pred. No. 3.3e-42;
 Matches 95; Conservative 27; Mismatches 41; Indels 4; Gaps 1;

OY 2 ALNSPLSGMGRIGRADFOCFQOARAVGLAGTFRATLSRLQDLYSIVRRADRAVPYV 61
 |||||
 DB 1222 ALNMFPSDIDR----ADQCKRQARAGLLSTYRFLSSRLQDLSTYRKAERYSLPYV 1277
 |||||

OY 62 LKDELFFPSWEALFSGSEGLPGARIFSPGKDVLRHPMPKSWHSGSDPGRRLTES 121
 |||||
 DB 1278 LKQVLFNPMDSIFSGHGQGMHMPHPIPSFGKDVLRHPMPKSWHSGSDPGRRLTES 1337
 |||||

OY 122 YCETWTEAPSATGQASSLLGRLGQASASCHHAYIVLCIENSFMT 168
 |||||
 DB 1338 YCEMTKADTAVTGLASPLSTGKILDKAKYSCANRLIVLCIENSFMT 1384
 |||||

RESULT 5
 T22002
 hypothetical protein F39H11.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence, revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22002
 R:White, S.
 Submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19500
 A:Accession: T22002
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-650 <WTL>
 A:Cross-references: EMBL:281079; PIDN:CA03084.1; GSPDB:GN00019; CESP:F39H11.4
 A:Experimental source: Clone F39H11
 C:Genetics:
 A:Gene: CESP:F39H11.4
 A:Map position: 1
 A:Insertions: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 40.0%; Score 357; DB 2; Length 650;
 Best Local Similarity 44.0%; Pred. No. 8.1e-28;
 Matches 73; Conservative 24; Mismatches 63; Indels 6; Gaps 4;

OY 1 VALNSPLSGMGRIGRADFOCFQOARAVGLAGTFRATLSRLQDLYSIVRRADRAVPYV 60
 |||||

DB 468 IALSPFSGNHLGLGADLCYREARACGTTTFPAMLSNNQDLVRIYHSD-EDTIVV 526
QY 61 NIKDELLEPPSWALEFSGSEGLPKPGARIFSFQKDVLRHPTVPKQSVHSGSDPNRRRLTE 120
DB 527 VAGAHHLFPPSWRMSFVNGAQ--WNPHAKLFSPDRHVDLNSRMPDKRVHSGSDGGR-AE 583
QY 121 SYCEWRTREAPSATQOASSLLGRLGSAAS--CHNAIVYCIEN 164
DB 584 QYCDGWRMRADSLTSLAGHISSTNTSIFSSGSEKCNKLVLCVEN 629

RESULT 6
T05202

pectinesterase homolog F4110.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C:Accession: T05202
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hohenseel, J.; Men
submitted to the Protein Sequence Database, February 1999
Reference number: 215402
Accession: T05202
Molecule type: DNA
A:Residues: 1-477 <BEV>
A:Cross-references: EMBL:AL035525
A:Experimental source: cultivar Columbia; BAC clone F4110
C:Genetics:
A:Map position: 4
A:introns: 7/3; 47/1; 90/1; 246/1; 354/1
A:Note: F4110.150
C:Superfamily: pectinesterase

Query Match 8.8%; Score 79; DB 2; Length 477;
Best Local Similarity 21.9%; Pred. No. 5.2;
Matches 32; Conservative 25; Mismatches 53; Indels 36; Gaps 6;
QY 3 LNSPLSGMGRGIRGADFCQOARAVGLAGTFRATLSRLDLYIVRADRAVPYINL 62
DB 63 LSAALSNQATCMEGRD-----GTSGLVKSLVAGSLDLYSMLE-----LLPIVQ- 107
QY 63 KDELLFPSEWALEFSGSEGLPKPGARIFSFQKDVLRHPTVPKQSVHSGSDPNRR 116
DB 108 -----PEQKPKAVSKPPGIAPKGPAPGRKLRDDEDESLQFPMVVR-----PDDR 153
QY 117 RUTESYCEWRTRE-APSATQOASSLL 141
DB 154 KLESNGRTYDVSVALDGTGNFTKIM 179

RESULT 7
3054

succinoglycan biosynthesis protein exom [imported] - Agrobacterium tumefaciens (strain C
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG3054
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
erage, G.; Gillet, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCiell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG3054
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-314 <KUR>
A:Cross-references: GB:AE008689; PIDN:ALA44853.1; PID:9177442499; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: exom
A:Map position: linear chromosome
C:Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

Query Match 8.6%; Score 77; DB 2; Length 314;
Best Local Similarity 27.9%; Pred. No. 5;
Matches 29; Conservative 14; Mismatches 41; Indels 20; Gaps 5;
QY 64 DELLEPPSW-----BALFSGSEGLPKPGARIFSFQKDVLRHPTVPKQSVHSGSDP---N 114
DB 105 DETAPPHMLTALLETAETGATVYCGPYTAVYRNA-----PGMKKGRDFHSTVPVWVN 158
QY 115 GRRLTESYCEWRTREAPSATQOASSLLGRLGSAASCHHAY 157
DB 159 GEITGYTCNTLLRMEAPAVKGRPAL-----ALGSGSGEDTHFF 198

RESULT 8
F98231

succinoglycan biosynthesis protein exom [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: F98231
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldm
A.; Liu, F.; Woliam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: AG7359; MUID:21608551; PMID:11743194
Accession: F98231
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-314 <KUR>
A:Cross-references: GB:AE007870; PIDN:NAK89376.1; PID:915159227; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1606
A:Map position: linear chromosome
C:Superfamily: Rhizobium succinoglycan biosynthesis glycosyltransferase

Query Match 8.6%; Score 77; DB 2; Length 314;
Best Local Similarity 27.9%; Pred. No. 5;
Matches 29; Conservative 14; Mismatches 41; Indels 20; Gaps 5;
QY 64 DELLEPPSW-----BALFSGSEGLPKPGARIFSFQKDVLRHPTVPKQSVHSGSDP---N 114
DB 105 DETAPPHMLTALLETAETGATVYCGPYTAVYRNA-----PGMKKGRDFHSTVPVWVN 158
QY 115 GRRLTESYCEWRTREAPSATQOASSLLGRLGSAASCHHAY 157
DB 159 GEITGYTCNTLLRMEAPAVKGRPAL-----ALGSGSGEDTHFF 198

RESULT 9
B45022

CRK-I - human
C:Species: Homo sapiens (man)
C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 12-Feb-1999
C:Accession: B45022
R:Matsumi, M.; Tanaka, S.; Nagata, S.; Kojima, A.; Kureta, T.; Shibuya, M.
Mol. Cell. Biol. 12, 3482-3489, 1992
A:Title: Two species of human CRK cDNA encode proteins with distinct biological activ
A:Reference number: A45022; MUID:92334347; PMID:1630456
A:Accession: B45022
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-204 <MAT>
A:Experimental source: embryonic lung cells
A:Note: sequence extracted from NCBI backbone (NCBI:108771, NCBI:108772)
C:Superfamily: crk transforming protein; SH2 homology; SH3 homology
F:13-118/Domain: SH2 homology <SH2>
F:139-187/Domain: SH3 homology <SH3>

Query Match 8.5%; Score 76; DB 2; Length 204;
Best Local Similarity 29.0%; Pred. No. 3.8;
Matches 20; Conservative 12; Mismatches 21; Indels 16; Gaps 3;
QY 88 IFSPDGK-----DVLRHPTVPKQSVHSGSDPNRR--LRESYCEWRTREAPSATG 135

Db 140 LEFENGNEEDLPFKKGLILRLROKPEQOMWNAEDSEGRKMIPIVYEKTR----PASA 195
QY 136 QASSLGR 144
Db 196 SVSALIGR 204

RESULT 10

Class I cytokinase receptor precursor - human
N:Alternate names: MSX-1
C:Species: Homo sapiens (man)
C>Date: 17-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 21-Jul-2000
C:Accession: J00047
R:Specier, C.A.; Grant, F.J.; Baumgartner, J.W.; Presnell, S.R.; Schrader, S.K.; Yamagi
Biochem. Biophys. Res. Commun. 246, 82-90, 1998
A:Title: Cloning and characterization of a novel class I cytokine receptor.
A:Reference number: J00047; MUID:98262921; PMID:9600072
A:Accession: J00047

A:Molecule type: mRNA
Residues: 1-636 <SPR>
Cross-references: GB:AF053004; NID:93153240; PIDN:AC39755.1; PID:93153241
Experimental source: brain
C:Genetics:

A:Map position: 19p13.11
C:Keywords: glycoprotein
F:1-32/Domain: signal sequence #status predicted <SIG>
F:515-540/Domain: transmembrane #status predicted <TM>
F:554-561/Domain: cytoplasmic #status predicted <CMP>
F:51,76,302,311,374,382,467/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match Best Local Similarity 8.5%; Score 76; DB 2; Length 636;
Matches 39; Conservative 12; Mismatches 46; Indels 64; Gaps 7;

QY 11 MRGIRGADF-----QCQQAARAVGLAGTFRALSSRLQDL----- 45
Db 1 MRGRGAPFWLWLPKLLALPLWLVLFOFRRRPGSSAGPLQCYGVGLDLNCSEWEPGLDL 60

QY 46 -----YSIVRRADRAAVPI-----NLKDELFE-----PSMEALFS 76
Db 61 GAPSELHLOSQKYSNKTQYVAAGRSWVAIPREQLTMSKLLVWGTCKAGQPLMPPEV 120

QY 77 GSEPLRPGA-RI---FSPDGKDLR-----HPTWPKSV 107
Db 121 NLETKMPNAPRLGPDVDFSEDDPLFATVHWAPPTWSHKV 161

RESULT 11

Dbable pullulanase (EC 3.2.1.41) - barley

N:Alternate names: pullulanase
C:Species: Hordeum vulgare (barley)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 03-Jun-2002
C:Accession: T04377
R:Lox, F.; Kristensen, M.; Plancho, V.; Leah, R.; Svendsen, I.; Svendsen, B.
Submitted to the EMBL Data Library, December 1997
A:Description: Isolation and characterization of starch debranching enzyme, limit dextrin
A:Reference number: 215320
A:Accession: T04377

A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-904 <LOK>
A:Cross-references: EMBL:AF022725; NID:92502057; PIDN:AAD04189.1; PID:92677837
A:Experimental source: cv. Igri1
C:Genetics:

A:Gene: HYLD99
A:Accession: 21/3; 70/1; 87/3; 124/1; 169/3; 235/2; 285/2; 305/3; 346/1; 376/2; 407/3; 439
C:Keywords: glycosidase; hydrolase

Query Match Best Local Similarity 8.5%; Score 75.5; DB 2; Length 904;
Matches 45; Conservative 19; Mismatches 67; Indels 69; Gaps 8;

QY 5 SPLSGMIRGADPOCFQARAVLAGT-----FRAFLSSRLQDLTYIVR----- 50
Db 62 SPSNG---GIGYDSKVLQDPESAGLPETVQKFFPFISSYRAFVKPSSVDVSLVKCOLY 118
QY 51 ----RADRAAVPIVNL-----KDELLFPSME-----ALRSG 77
Db 119 VASFGADGKHVDVGLQLPGVLDWFAVTGFLGAVFSDSVSLHLMAPTAGVSYCFPDG 178
QY 78 SEGR-----LKPGRITSPDG-----DVLRHPTWPKSVWHSDDPGRRLT 119
Db 179 PAGPALETVOLKESNGVSVTGPREMENRYLYEVDVY-HPTKQVCLKLAGDPYARLS 237
QY 120 ESYCTWRTAPSATQASS 139
Db 238 ANGARTWLVDINNETLKPAS 257

RESULT 12

B89781

conserved hypothetical protein SA0184 [imported] - Staphylococcus aureus (strain N315
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B89781

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-O, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratazu, K.
Lancet 357, 1255-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: B89758; MUID:21311952; PMID:11418146

A:Accession: B89781
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-351 <KUR>
A:Cross-references: GB:BA000018; PID:913700106; PIDN:BA041405.1; GSPDB:GM00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0184

Query Match Best Local Similarity 8.4%; Score 75; DB 2; Length 351;
Matches 42; Conservative 17; Mismatches 67; Indels 54; Gaps 8;

QY 11 MRGIRGADPOCFQARAVLAGTFRALSSRLQDL-----YSIVRRADRA-AVPIV 60
Db 97 IEATMAOGLKCCCLNASTIS-----RELTTSLHQLNDFTLTSFCHNRYPPDGTGLSDVLY 151

QY 61 NLKDELFPSEWALFSGSEGPLKPGARITFSPDGKDLRHPPTWPKSVWHSDDPGRRLTE 120
Db 152 NKNKELLY-----QFNPKQITGFIYSGSLRGPL-----HKGLPT----- 186

QY 121 SYCTWRTAP-----SATGASSLLGRLGQSNAAS-----CHNAVTLCTIENSFMT 168
Db 187 --IEATRHSHPVAAKLLQETGVSEVLVGSLLIEMRQAKLIDPCRKIRHHTTLCIEEYFD 244

RESULT 13

S56015

gastric mucin MUC5AC - human (fragment)

C:Species: Homo sapiens (man)
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Apr-2000
C:Accession: S56015; S53361
R:Klamp, L.W.J.; van Rens, L.; Strous, G.J.
Biochem. J. 308, 831-838, 1995

A:Title: Cloning and analysis of human gastric mucin cDNA reveals two types of conser
A:Reference number: S56015; MUID:97104281; PMID:8948439
A:Accession: S56015

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-850 <KLO>
A:Cross-references: EMBL:X81649; NID:9547516; PIDN:CA57309.1; PID:9547517
R:Gyomai-Dupercat, V.; Audie, J.P.; Debailieu, V.; Latine, A.; Bulstine, M.P.; Galleg
Biochem. J. 305, 211-219, 1995

A:Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich domain
A:Reference number: S53361; MUID:95126907; PMID:7826332
A:Accession: S53361
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 648-678, 'L', 680-733, 'L', 735-760 <GUY>
A:Cross-References: EMBL:Z34280; NID:g563380; PIDN:CAA84034.1; PID:g563381
A:Experimental source: clone JUI32
A>Note: this publication is not cited in GenBank entry HSMUCIN5, release 113.0

Query Match 8.3%; Score 74.5; DB 2; Length 850;
Best Local Similarity 24.7%; Pred. No. 30;

Matches 39; Conservative 23; Mismatches 71; Indels 25; Gaps 6;

OY 13 GARGADFCQOARAVGLAGTFRALSRLODLYSIVRADR-AAVPIVNLKDELLFSPSW 71
DB 592 GINGDGFDFQNLREBQ--TF-----CESPRVQCAESEPNTPLADQDVCISHT 642
OY 72 EALFSGSECPKLP-----GARISFDGKDVLRHPTWPKSV-----HGSDPNGRRLTE 120
DB 643 ECLILINKNQLPPICYNEIRIQCCETVAVCRDITRPRKTVATTRPTGAGQQTFT 702
OY 121 SYCETWREAPSATQ-----ASSLGRLLGSAASCH 154
DB 703 THMPSASTEQPTATSRGGPTATSVQGHHTPTVRNCH 740

RESULT 14

B91052

hypothetical protein EC93386 [imported] - Escherichia coli (strain O157:H7, substrain R1)

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: B91052
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gesawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: B91052

A:Molecule type: DNA

A>Status: preliminary

A:Residues: 1-1653 <HAY>

A:Cross-References: GB:BA000007; PIDN:BA836809.1; PID:g13362856; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: EC93386

Query Match 8.2%; Score 73.5; DB 2; Length 1653;
Best Local Similarity 24.4%; Pred. No. 84;

Matches 32; Conservative 21; Mismatches 47; Indels 31; Gaps 5;

OY 29 GIAGTFRAFLSSRLDLY-----SIVRRADRAVPIVNLKDELLFSPMEALFSGSE 79
DB 380 GAPGYSKQFFMFGPRDLRPGETVILNGLRDADGKALPNQPIKIDYIKPDGQVLRYSVS 439
OY 80 GPLKPGARLFSFDGKDVLRHPTWPKS-----VWH-----GSDPNGRRLTESYCETWRT-- 129
DB 440 QP-----ENGLYHFTWPLDSNAATGMHIRANTGDNQYRMWDFHVEDFMPBRM 487
OY 130 APSATGQASSL 140
DB 488 ALNLTGKKTPL 498

RESULT 15

F85896

hypothetical protein Z3787 [imported] - Escherichia coli (strain O157:H7, substrain EDLg)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: F85896

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
1116, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85896
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1653 <STO>
A:Cross-References: GB:AE005174; NID:g12516921; PIDN:AGS7634.1; GSPDB:GN00145; UNMCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:

A:Gene: Z3787

Query Match 8.2%; Score 73.5; DB 2; Length 1653;
Best Local Similarity 24.4%; Pred. No. 84;

Matches 32; Conservative 21; Mismatches 47; Indels 31; Gaps 5;

OY 29 GIAGTFRAFLSSRLDLY-----SIVRRADRAVPIVNLKDELLFSPMEALFSGSE 79
DB 380 GAPGYSKQFFMFGPRDLRPGETVILNGLRDADGKALPNQPIKIDYIKPDGQVLRYSVS 439
OY 80 GPLKPGARLFSFDGKDVLRHPTWPKS-----VWH-----GSDPNGRRLTESYCETWRT-- 129
DB 440 QP-----ENGLYHFTWPLDSNAATGMHIRANTGDNQYRMWDFHVEDFMPBRM 487
OY 130 APSATGQASSL 140
DB 488 ALNLTGKKTPL 498

Search completed: July 24, 2003, 17:09:02
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:04:52 ; Search time 29 Seconds

(without alignments)
248.029 Million cell updates/sec

Title: US-09-171-607A-1

Perfect score: 893
Sequence: 1 VALNSPLSGMGRGIRGADFO.....ASCHHAYIVLCIENSFMTAS 170

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents.AA.*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep.*
7: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	893	100.0	178	4	US-09-315-689-5
2	893	100.0	182	4	US-09-561-500-14
3	893	100.0	182	4	US-09-561-108-14
4	893	100.0	182	4	US-09-315-689-3
5	893	100.0	182	4	US-09-561-526-14
6	893	100.0	182	4	US-09-561-499-14
7	893	100.0	183	3	US-09-306-059-2
8	778	87.1	191	4	US-09-561-500-13
9	778	87.1	191	4	US-09-561-108-13
10	778	87.1	191	4	US-09-561-526-13
11	778	87.1	191	4	US-09-561-499-13
12	775	86.8	195	1	US-08-159-784-2
13	734	82.2	185	3	US-08-985-526-36
14	491	55.0	191	1	US-08-159-784-3
15	160	17.9	35	3	US-09-046-985-2
16	160	17.9	35	3	US-09-474-743-2
17	101	11.3	22	3	US-09-046-985-7
18	101	11.3	22	3	US-09-474-743-7
19	94	10.5	16	3	US-09-385-442-32
20	76	8.5	403	4	US-09-252-991A-22238
21	76	8.5	578	1	US-08-653-740-3
22	76	8.5	578	2	US-09-073-594-3
23	76	8.5	578	3	US-09-275-925-3
24	76	8.5	636	1	US-08-653-740-5
25	76	8.5	636	2	US-09-073-594-5
26	76	8.5	636	3	US-09-275-925-5
27	74	8.3	256	1	US-07-906-349A-8

28	74	8.3	256	1	US-08-167-035-4	Sequence 4, Appl1
29	74	8.3	256	1	US-08-208-887A-4	Sequence 4, Appl1
30	74	8.3	256	2	US-08-539-005-4	Sequence 4, Appl1
31	74	8.3	256	4	US-09-280-598-8	Sequence 8, Appl1
32	73	8.2	311	3	US-08-987-743-2	Sequence 2, Appl1
33	73	8.2	435	3	US-08-733-360A-1	Sequence 1, Appl1
34	73	8.2	435	3	US-08-733-360A-3	Sequence 3, Appl1
35	73	8.2	435	3	US-08-987-743-6	Sequence 6, Appl1
36	73	8.2	435	3	US-08-987-743-15	Sequence 15, Appl1
37	73	8.2	435	3	US-08-916-935-3	Sequence 3, Appl1
38	73	8.2	435	3	US-08-916-935-3	Sequence 19779, A
39	72.5	8.1	184	4	US-09-252-991A-19779	Sequence 28, Appl1
40	72	8.1	304	4	US-08-630-915A-28	Sequence 12, Appl1
41	72	8.1	1006	4	US-09-023-905A-12	Sequence 6980, Ap
42	70.5	7.9	337	4	US-09-328-352-6980	Sequence 29674, A
43	70.5	7.9	409	4	US-09-252-991A-29674	Sequence 148, App
44	70	7.8	210	3	US-09-188-930-148	Sequence 148, App
45	70	7.8	210	4	US-09-312-283C-148	Sequence 148, App

ALIGNMENTS

RESULT 1
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkmann, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antitumor Endostatin Compositions
; FILE REFERENCE: 03213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5

Query Match 100.0%; Score 893; DB 4; Length 178;
Best Local Similarity 100.0%; Pred. No. 3.9e+101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	VALNSPLSGMGRGIRGADFOCFQARAVGLAGTFRATLSRLDLYSIVRRADRAVPIV	60
DB	9	VALNSPLSGMGRGIRGADFOCFQARAVGLAGTFRATLSRLDLYSIVRRADRAVPIV	68
QY	61	NLNDLFPMEALFSGEGPLKPGARIFSGDKDVRHPTWPKSWHSGSDPGRRLTE	120
DB	69	NLNDLFPMEALFSGEGPLKPGARIFSGDKDVRHPTWPKSWHSGSDPGRRLTE	128
QY	121	SYCTWTEAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS	170
DB	129	SYCTWTEAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS	178

RESULT 2
US-09-561-500-14
; Sequence 14, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-14

Query Match 100.0%; Score 893; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 4.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 72
OY 61 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 120
DB 73 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 132
OY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 3

US-09-561-108-14
Sequence 14, Application US/09561108
Patent No. 6342221

GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Breken
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/561.108
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-14

Query Match 100.0%; Score 893; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 4.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 72
OY 61 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 120
DB 73 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 132
OY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 4

US-09-315-689-3
Sequence 3, Application US/09315689
Patent No. 6346510
GENERAL INFORMATION:

APPLICANT: Folkman, Judah
APPLICANT: O'Reilly, Michael
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
FILE REFERENCE: 05213-0229
CURRENT APPLICATION NUMBER: US/09/315.689
CURRENT FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 182
TYPE: PRT
ORGANISM: Homo sapiens
US-09-315-689-3

Query Match 100.0%; Score 893; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 4.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 72
OY 61 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 120
DB 73 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 132
OY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 5

US-09-561-526-14
Sequence 14, Application US/09561526
Patent No. 6416758

GENERAL INFORMATION:
APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Breken
TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002586
CURRENT APPLICATION NUMBER: US/09/561.526
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-14

Query Match 100.0%; Score 893; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 4.1e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRALFSSRLQDLYSIVRRADRAAIVY 72
OY 61 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 120
DB 73 NLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHGSDPNGRRLTE 132
OY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 6

Db 20 VALNTPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADGSPVIV 79
Qy 61 NLKDELLPSPWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVWHGSDPNGRRLTE 120
Db 80 NLKDEVLSPSWDSLFSSGQGLQOPGARIFSPDGRVLRHPMPQKSVWHGSDPNGRRLME 139
Qy 121 SYCETWRTETGATGQASSLLSGRLLEGQSAASHHAYIVLCIENSFMTA 169
Db 140 SYCETWRTETGATGQASSLLSGRLLEGQSAASHHAYIVLCIENSFMTS 188

RESULT 10
US-09-561-526-13
; Sequence 13, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-13

Query Match 87.1%; Score 778; DB 4; Length 191;
Best Local Similarity 85.8%; Pred. No. 4.9e-87;
Matches 145; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 1 VALNSPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADRAAPVIV 60
Db 20 VALNTPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADGSPVIV 79
Qy 61 NLKDELLPSPWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVWHGSDPNGRRLTE 120
Db 80 NLKDEVLSPSWDSLFSSGQGLQOPGARIFSPDGRVLRHPMPQKSVWHGSDPNGRRLME 139
Qy 121 SYCETWRTETGATGQASSLLSGRLLEGQSAASHHAYIVLCIENSFMTA 169
Db 140 SYCETWRTETGATGQASSLLSGRLLEGQSAASHHAYIVLCIENSFMTS 188

RESULT 11
US-09-561-499-13
; Sequence 13, Application US/09561499
; Patent No. 6524583
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match 87.1%; Score 778; DB 4; Length 191;
Best Local Similarity 85.8%; Pred. No. 4.9e-87;
Matches 145; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

Qy 1 VALNSPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADRAAPVIV 60
Db 20 VALNTPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADGSPVIV 79
Qy 61 NLKDELLPSPWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVWHGSDPNGRRLTE 120
Db 80 NLKDEVLSPSWDSLFSSGQGLQOPGARIFSPDGRVLRHPMPQKSVWHGSDPNGRRLME 139
Qy 121 SYCETWRTETGATGQASSLLSGRLLEGQSAASHHAYIVLCIENSFMTA 169
Db 140 SYCETWRTETGATGQASSLLSGRLLEGQSAASHHAYIVLCIENSFMTS 188

RESULT 12
US-08-159-784-2
; Sequence 2, Application US/08159784
; Patent No. 5643783
; GENERAL INFORMATION:
; APPLICANT: Bjorn R. Olsen
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 86.8%; Score 775; DB 1; Length 195;
Best Local Similarity 85.2%; Pred. No. 1.2e-86;
Matches 144; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 VALNSPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADRAAPVIV 60
Db 24 VALNTPLSGMRGIRGADFOCFQOARAVGLSTFRAPFLSSRLQDLXSVIRRADGSPVIV 83
Qy 61 NLKDELLPSPWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVWHGSDPNGRRLTE 120
Db 84 NLKDEVLSPSWDSLFSSGQGLQOPGARIFSPDGRVLRHPMPQKSVWHGSDPNGRRLME 143

OY 121 SYCETWRTAPATGQASSLLGGRLGQASCHAHYIVLCIENSPMTA 169
DB 144 SKCETWRTETGATGQASSLLSGRLLEOKAASCHNSYIVLCIENSPMTS 192

RESULT 13

US-08-985-526-36
Sequence 36, Application US/08985526
Patent No. 6080728
GENERAL INFORMATION:
APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
NUMBER OF INVENTIONS: 43
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: McMorrow Jr., Robert G
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-36

Query Match 82.2%; Score 734; DB 3; Length 185;
Best Local Similarity 82.4%; Pred. No. 1.1e-81;
Matches 140; Conservative 14; Mismatches 14; Indels 2; Gaps 2;

OY 1 VALNSPLSGMGRIGADFCQFOARAVAGLAFRAFLSSRLDLYSIVRADAAVPIV 60
DB 14 VALNTPLSGMGRIGADFCQFOCFNNAR-VGLSGTFRALSSRLDLYSIVRADAGSVPIV 72
OY 61 -NLKDELLFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPPTWOKSVHSGDPNGRRLT 119
DB 73 QNFRDEVLSPSMWLSFGSQGQLQPGARIFSPDGKDVLRHPPTWOKSVHSGDPNGRRLM 132
OY 120 ESYCEWRTAPATGQASSLLGGRLGQASCHAHYIVLCIENSPMTA 169
DB 133 ESYCEWRTETGATGQASSLLSGRLLEOKAASCHNSYIVLCIENSPMTS 182

RESULT 14

US-08-159-784-3
Sequence 3, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 330
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 191
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-3

Query Match 55.0%; Score 491; DB 1; Length 191;
Best Local Similarity 56.3%; Pred. No. 5.8e-52;
Matches 94; Conservative 27; Mismatches 42; Indels 4; Gaps 1;

OY 2 ALNSPLSGMGRIGADFCQFOARAVAGLAFRAFLSSRLDLYSIVRADAAVPIV 61
DB 25 ALNPFSGDIR----ADRCFKQARAAGLLSTYRAPLSSHQDLSITYRKERYSLPIVN 80
OY 62 LKDELLFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPPTWOKSVHSGDPNGRRLTES 121
DB 81 LKGVLEFNMDISFSGHGGQGPNNHPIYFSGDRIMTDPSPQKVIYHMGSPHGVRLVDN 140
OY 122 YCETWRTAPATGQASSLLGGRLGQASCHAHYIVLCIENSPMT 168
DB 141 YCEAMRTADYAVTGLASPLSTGKTLIDOKAVSCANRLIYLCIENSPMT 187

RESULT 15

US-09-046-985-2
Sequence 2, Application US/09046985
Patent No. 6121236
GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,985

FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: BROOK, David E.
 REGISTRATION NUMBER: 22,592
 REFERENCE/DOCKET NUMBER: CMCC-614
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781) 861-6240
 TELEFAX: (781) 861-9540
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 35 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-09-046-985-2

Query Match 17.9%; Score 160; DB 3; Length 35;
 Best Local Similarity 100.0%; Pred. No. 1.4e-12;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QARAVGLAGTFRAPFLSSRLQDLYSTVRRADRAAY 57
 DB 1 QARAVGLAGTFRAPFLSSRLQDLYSTVRRADRAAY 34

Search completed: July 24, 2003, 17:09:38
 Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:16:08 ; Search time 40 Seconds
(without alignments)
408.717 Million cell updates/sec

Title: US-09-171-607a-1

Perfect score: 893
Sequence: 1 VALNSPLSGMGKRGADFO.....ASCHRAYIVLCIENSFMPTAS 170

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 85130

Minimum DB seq length: 0
Maximum DB seq length: 170

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.5	7.8	168	2	S72898
2	67	7.5	145	1	J00947
3	64	7.2	126	2	A69177
4	63.5	7.1	149	2	A46132
5	63	7.1	113	2	C95341
6	62.5	7.0	146	2	B96540
7	62	6.9	158	2	S61212
8	60.5	6.8	93	2	JF0055
9	60.5	6.8	156	2	A87366
10	60	6.7	145	1	ZWECAP
11	60	6.7	151	2	D84296
12	60	6.7	152	2	B72690
13	59.5	6.6	74	2	S72624
14	59	6.6	83	2	G84100
15	59	6.6	150	2	A11968
16	58.5	6.6	83	2	S7058
17	58.5	6.6	126	2	S77058
18	58.5	6.6	134	2	B7237
19	58.5	6.6	168	2	D75259
20	57	6.4	107	2	AF1948
21	57	6.4	127	2	T35336
22	57	6.4	159	2	T44144
23	56.5	6.3	82	2	AD3161
24	56.5	6.3	105	2	G01976
25	56.5	6.3	133	2	S04343
26	56.5	6.3	140	2	F70515
27	56	6.3	139	2	P00062
28	56	6.3	144	2	T43499
29	55.5	6.2	159	2	G75277

30	55	6.2	87	2	S31247	endozepine - yeast
31	55	6.2	91	2	I57611	MHC K-lme transp
32	55	6.2	107	2	F71121	hypothetical prote
33	54.5	6.1	165	2	G72511	hypothetical prote
34	54	6.0	115	2	AE3568	integral membrane
35	54	6.0	137	2	I80172	class I histocompa
36	54	6.0	162	2	D83332	toluate 1,2-dioxy
37	54	6.0	164	2	T36486	hypothetical prote
38	54	6.0	168	2	E70567	hypothetical prote
39	54	6.0	169	1	H69136	hypothetical prote
40	53.5	6.0	91	2	G01975	calmodulin-depende
41	53.5	6.0	101	2	AD3542	hypothetical prote
42	53.5	6.0	136	1	R5K716	hypothetical prote
43	53.5	6.0	137	2	I38875	ribosomal protein
44	53.5	6.0	146	2	S06775	MHC class I anti
45	53.5	6.0	157	2	T37041	much protein - Sal
						hypothetical prote

ALIGNMENTS

RESULT 1
S72898
hypothetical protein B2168_C2_214 - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S72898
R:Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid B2168.
A:Reference number: S72586
A:Accession: S72898
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-168 <SM1>
A:Cross-references: EMBL:U00018; NID:g467037; PIDN:AA17234.1; PID:g467050

Query Match
Best Local Similarity 29.5%; Pred. No. 13;
Matches 36; Conservative 11; Mismatches 44; Indels 31; Gaps 6;

QY 3 LNSPLSGMGKRGADFOCFQOARAVGL-----AGTFRAPLSSRLDOLXIVRAD 53
DB 14 LAGPLVPMCG-RDARRDLLEODNLGALBCAAMTGKGTN-IGASGIMLSQAIRAG 71
QY 54 RAAVPI-----VNLKDELFPSEALFSGSEGPLKPGARIFSGDKVLRHP 100
DB 72 RIAPPIPSFGVWALDLRWVRYREITRDQFYL---SYGNVMDTTRM---GSELGYHP 124
QY 101 TW 102
DB 125 KW 126

RESULT 2
J00947
atrial natriuretic peptide precursor - bullfrog
N:Alternate names: ANP; atrial natriuretic factor (ANF)
M:Contents: atrial natriuretic peptide-21; atrial natriuretic peptide-24
C:Species: Rana catesbeiana (bullfrog)
C:Date: 10-Sep-1989 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
R:Kojima, M.
submitted to JIPID, May 1991
A:Reference number: J00947
A:Accession: J00947
A:Molecule type: mRNA
A:Residues: 1-145 <KOJ>
R:Sakata, J.; Kangawa, K.; Matsuo, H.
Biochem. Biophys. Res. Commun. 155, 1338-1345, 1988
A:Title: Identification of new atrial natriuretic peptides in frog heart.
A:Reference number: A31510; MUID:89025806; PMID:2972279
A:Accession: A31510

A:Molecule type: protein
A:Residues: 132-145 <SAK>
C:Comment: In mammals, several active peptides may be derived from the prohormone
C:Superfamily: natriuretic peptide A precursor
C:Keywords: atrium; diuretic; hormone; natriuretic; osmoregulation
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-121/Domain: propeptide #status predicted <PRO>
F:132-145/Product: atrial natriuretic peptide-24 #status experimental <M24>
F:135-145/Product: atrial natriuretic peptide-21 #status experimental <M21>
F:135-141/Disulfide bonds: #status predicted

Query Match	7.5%	Score 67;	DB 1;	Length 145;
Best Local Similarity	21.1%;	Pred. No. 20;		
Matches 34;	Conservative 24;	Mismatches 49;	Indels 54;	Gaps 9;

QY 2 ALNSPLSGMGKRGIRADFOCFQQAARVGLACTFRALFLSSRLDLYSLVRADDAAPLYVN 61
| |||| : ::| | ||| : :|
Db 26 AYNSPPLSDSLDLK-----LLERIEDRLPV----EEVETPV-- 58

Db

```
62 LKIDLLFSPMEALFSGSECF-----LKPGARIFSDGKDVLRHPWPKQSWHGSDPNG 115
   :: ::: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
59 --QDIFAPNVDSADSSNSAPSLTYEARPGADMVNRG-----SWTQQ---EKSSPLR 106
```

```

QY      116 RLTESYCETWRTAPSATGQASSLLGGRL--LG-QSASC 153
      :| | : :| | : :| | |
Db      107 NKLRE-----LLNAPRSMRRSSDCFGSRIDRIGAQSGMGC 141

```

RESULT 3
A69177
hypothetical protein MTH581 - Methanobacterium thermoautotrophicum (strain Delta H)

R:Smith, P.R.; Doucette-Stamm, J.A.; Deloughery, C.; Lee, H.; Dubois, J.; A
C:Species: *Methanobacterium thermoautotrophicum*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: A69177

Qin, D., Spauldrew, K., Vicalle, R., Wang, I., Wietzowski, J., Gibson, R., Jiwani, N. K. S., Church, G. M., Daniels, C. J., Mao, J., Rice, P., Noelling, J., Reeve, J. N. *J. Bacteriol.* 179, 7135-7155, 1997

A:Reference number: A05000; MOLID:9803314; PMID:9314063
A:Accession: A69177
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
translation not shown

A:Accession: J120-1474
A:Cross-references: GB:AE000840; GB:AE000666; MID:g2621653; PIDN:AAB85087.1; PID:g2621653
A:Experimental source: strain Delta H
C:Genetics:

Query Match 7.2% Score 64 DB 2: length 126:
Methanobacterium thermoautotrophicum hypothetical protein MTH581

Category	Count	Percentage
Matches	21	10%
Mismatches	10	5%
Indels	30	15%
Gaps	4	2%

Db . 17 AELSDLRLFEKKANLHGACVKKLLTDNLSFKKEENANFKMSDDVYRPAKYFAISDGDLE 76

Oy 96 VLRRHP 100

Db 77 VLYEP 81

A46132
c-Jun leucine zipper interactive (CDNA JZA-3) - mouse (fragment)
C1Species: Mus musculus (house mouse)

C:Accession: A46322
R:Chevrey, P.M.; Nathans, D.
Proc. Natl. Acad. Sci. U.S.A. 89, 5789-5793, 1992

A: Reference number: A46132; MUID:9235183; PMID:1631061

A:Accession: A46132
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-149 <GHE>
A:Experimental source: embryo
A:Note: sequence extracted from NCBI backbone (NCBIP:108531)

Query Match	7.1%	Score 63.5;	DB 2;	Length 149;
Best Local Similarity	24.7%	Pred. No. 47;		
Matches 20;	Conservative 11;	Mismatches 45;	Indels 5;	Gaps 1.

Qy 44 DLYSIVRRDRAAVPIVNLKDELFFSWEALFSGSEGPLKRGARISFDGKQVLRHPIWP 103

Db 55 DLQLAQEDAKKMAV----KEEKYDPGYEAAGAYGENPCNGEPCGIASNGLTAAHSAEP 109

```

QY      104 QKSVHGSDDPNGRRLTESYCE 124
          :      |      ||| : : : |
DB      110 RGEATPGDVPNGQWMAQSFAP 130

```

RESULT 5

C95341 C95340 (continued) C95340 (continued)

C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: C95341

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A:Reference number: A95262; MUID:21386505; PMID:11481432

A:Accession: C95341
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-113 <KIR>

Alcious-terferences: GR:NEU040637; ELOD:AA603235.1; ELOD:914323746; GSFDB:G000430
A:Cross-references: GR:NEU040637; ELOD:AA603235.1; ELOD:914323746; GSFDB:G000430
Experimental source: strain 1021, megaplasmid pSYMA
R:Calberlet, F.; Flann, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barley-Hubb
Bela, D.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federfeld, N.A.; Fisher, R.
Bela, D.; Chait, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federfeld, N.A.; Fisher, R.

L.: Hyman, R.-W.; Jones, I. Science 293, 668-672, 2001.

A: Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelatu nebutau, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, P.

A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:

Query Match

Sequence	Best Local Similarity	Field No.	Matches	Conservative	Mismatches	Indels	Gaps
64 DELLPSWEALFSGSEGLPKPGAR-----IFS-----FDGKDLVLRHPTTPQKSVWHGS	27.44%	20	23	11	22	28	4

```

Db      45 DEIVVTKEALLDIAEPPLCNECRLOQYIAVFSDIASIKFDGHELA----- 90
Oy      112 DPNGR-RLTESYCEIWTREAPSAT 134

```

Db 91 - PDGRVAVTAADSVWVKVNHPEAT 113

hypothetical protein FL413.3 (Imported) - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

C:\Access\B96540
R:\Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
C:\Access\B96540

Nature 408, 816-820, 2000

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:09:43 ; Search time 22 Seconds

(without alignments)
363.388 Million cell updates/sec

Title: US-09-171-607A-1

Perfect score: 893
Sequence: 1 VALNSPLSGMRGIRGADFO.....ASCHHAYIVLCIENSFMPTAS 170

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 33346

Minimum DB seq length: 0
Maximum DB seq length: 170

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	67	7.5	145	1 ANF_RANCA	P18909 rana catesb
2	63.5	7.1	133	1 NMTL_DROME	O9V3H8 drosophila
3	59.5	6.7	174	1 SOML_YEAST	O05676 saccharomyc
4	56	6.3	146	1 HBB2_ANAMI	P83273 ananichas
5	55.5	6.2	159	1 Y006_DEIRA	O9VRS9 delinococcus
6	55	6.2	86	1 ACBP_YEAST	P31787 saccharomyc
7	54	6.0	144	1 ACB2_SACPS	O32272 saccharomyc
8	54	6.0	136	1 MERC_THIFE	P232905 thlobacilli
9	53.5	6.0	146	1 RK16_CVAPA	P23406 cyanophora
10	53.5	6.0	145	1 MUCA_SALTY	P07376 salimonia
11	53.5	6.0	153	1 C554_RHOSH	O53142 rhodosphe
12	53	5.9	128	1 CYCP_RHOSP	P00146 rhodosphe
13	53	5.9	140	1 HBAD_COLLI	O12983 rhodosphe
14	53	5.9	165	1 UBCJ_HUMAN	P26554 homo sapien
15	53	5.9	168	1 UBCJ_HUMAN	O00102 schizosacch
16	52.5	5.9	138	1 GSP1_XANCP	P31738 xanthomonas
17	52	5.8	90	1 YMN3_OENBE	P18633 cenothea b
18	52	5.8	158	1 IF53_ARATH	O9C505 arabidopsis
19	52	5.8	166	1 2SS4_ARATH	P15460 arabidopsis
20	51.5	5.8	109	1 CYC_TETRY	P00079 tetrahymena
21	51.5	5.8	122	1 NTF2_ORISA	O9VJ54 oryza sativ
22	51.5	5.8	167	1 UBCF_SCHPO	O9Y818 schizosacch
23	51	5.7	125	1 VG61_BPMU5	O05274 mycobacteri
24	50.5	5.7	134	1 RS9_THEMA	O9X184 thermotoga
25	50.5	5.7	143	1 IF2B_METUA	O57362 methanococc
26	50.5	5.7	164	1 UBC7_CABEL	P34477 caenorhabdi
27	50	5.6	104	1 VAFI_THETH	P74903 thermus the
28	50	5.6	107	1 CURE_STRCN	O02588 streptomyce
29	50	5.6	119	1 VP26_HSEVB	P28974 equine hearp
30	50	5.6	142	1 HBAA_SEROU	O9P9M4 seriola qui
31	50	5.6	150	1 RL15_THETH	P74910 thermus the
32	50	5.6	156	1 PM20_SCHPO	O14313 schizosacch
33	50	5.6	168	1 MLRS_RAV	P04466 rattus norv

34	49.5	5.5	75	1 REGB_PSEAE	O03381 pseudomonas
35	49.5	5.5	113	1 LACT_AGRRD	O01936 agrobacteri
36	49.5	5.5	123	1 ADSV_PIG	Q29297 sus scrofa
37	49.5	5.5	138	1 RS8_THETH	P24319 thermus the
38	49.5	5.5	143	1 YW64_MYCTU	P96889 mycobacteri
39	49.5	5.5	168	1 RS8_AOUAE	O67566 aquifex aco
40	49	5.5	114	1 YPLI_DROME	O9W227 drosophila
41	49	5.5	124	1 MYL1_MOUSE	O88653 mus musculu
42	49	5.5	124	1 VEA_HPV57	P22157 human papil
43	49	5.5	128	1 ANF_CAVPO	P27596 cavia porce
44	49	5.5	137	1 ALFC_CHICK	P53449 gallus gall
45	49	5.5	142	1 HBAC_CONCO	P83479 conger cong

ALIGNMENTS

RESULT 1
ID ANF_RANCA STANDARD: PRT: 145 AA.
AC P18909;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Atrial natriuretic factor precursor (ANF) (Atrial natriuretic peptide)
DE (ANF).
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranidae; Rana.
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RA Kojima M.;
RL Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE=Heart.
RC MEDLINE=89025606; PubMed=2972279;
RX Sakata J., Kangawa K., Matsuo H.;
RT "Identification of new atrial natriuretic peptides in frog heart."
RL Biochem. Biophys. Res. Commun. 155:138-145(1988).
CC -1- FUNCTION: VASOACTIVE ACTIVITY. HAS A CGMP-STIMULATING ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE NATRIURETIC PEPTIDES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: D01043; BAA00850.1; -
DR PIR: JQ0947; JQ0947;
DR InterPro: IPR002407; At_natriureticp.
DR InterPro: IPR000663; Natr_peptide.
DR Pfam: PF00212; ANP. 1.
DR PRINTS: PRO0710; NATPEPTIDES.
DR Prodom: PD005107; At_natriureticp.
DR SMART: SM00183; NAT_PEP. 1.
DR PROSITE: PS00263; NATRIURETIC_PEPIDE; 1.
KW vasoactive; signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 119
FT PEPTIDE 122 145 ATRIAL NATRIURETIC FACTOR.
FT DISUFID 125 141
SQ SEQUENCE 145 AA: 15934 MW: 01EB3B9779529405 CRC64;

Query Match 7.5%; Score 67; DB 1; Length 145;
Best Local Similarity 21.1%; Pred. No. 11;
Matches 34; Conservative 24; Mismatches 49; Indels 54; Gaps 9;

```

OY 2 ALNSPLSGMGRICGADFCQOARAVAGLAGTFAFLSSRLDLYSIYRRADRAAPVYN 61
DB 26 AYNSPLSSDLSLKLK-----LLERLDRDPV-----EEVEPV-- 58
OY 62 LKDELFPMSWEALFSGSGSP-----LKPGARIFSDCKDVLRRHTWPOKSWHNSDPNG 115
DB 59 --QDFAFYDSDADSSNSAPSLTVEPARGADMMYNRG-----SWTQO--EKSSPLR 106
OY 116 RLRTSECYETWRTAPSATGQASLLGRL--LG-QSASA 153
DB 107 NKLRE-----LLNAPRSMRNSDCEGSRIDRIGASGMC 141

RESULT 2
ID NXT1_DROME STANDARD: PRT: 133 AA.
AC 09V3H8:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
RT NTF2-related export protein (p15).
RT Mol. Cell. Biol. 19:8616-8624(1999).
RT [1]
RX MEDLINE=20036817; PubMed=10567585;
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;
RT Identification of an NTF2-related factor that binds Ran-GTP and
RT regulates nuclear protein export.
RT Mol. Cell. Biol. 19:8616-8624(1999).
RT [2]
RX SEQUENCE FROM N.A.
RA STRAIN-Berkeley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKlos G.L.G.,
RA Abriell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Clawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrar C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazuelo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenklamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletter E., Spradling A.C., Stapleton M., Strong R.J., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zavelle J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

```

RA Glibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
CC -1- FUNCTION: Stimulator of protein export for NES-containing
CC proteins. Also plays a role in the nuclear export of. U1 snRNA,
CC tRNA, and mRNA (By similarity).
CC -1- SUBUNIT: Preferentially binds Ran-GTP (By similarity). Associates
CC with NXF1.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Contains 1 NTF2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF156959; AAD54944.1; -
DR EMBL: AE003462; AAF47066.1; -
DR FlyBase: FBgn0028411; Nxt1.
DR GO: GO:0016973; P:poly(A)+ mRNA-nucleus export; IMP.
DR InterPro: IPR002073; NTF2.
DR Pfam: PF02136; NTF2; 1.
DR PROSITE: PS50177; NTF2_DOMAIN; 1.
DR Transprot; Protein transport; mRNA transport; Nuclear protein.
KW DOMAIN 15 130
FT SEQUENCE 133 AA; 15181 MW; C6E664950AA370AA CRC64;
SQ
Query Match: 7.1%; Score 63.5; DB 1; Length 133;
Best Local Similarity 32.1%; Pred. NO. 21;
Matches 17; Conservative 13; Mismatches 18; Indels 5; Gaps 2;
OY 108 WHGSDPNRRLTESCYETWRTAPSATGQASLLGRLGSAASCHAYIVL 160
DB 46 WNGNGAIGROMESTFYQ---ELPSSNHQLNLDAQPIVDO-AVSNQIAYILM 93
ID NXT1_DROME STANDARD: PRT: 74 AA.
AC 005676;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
RT SOM1 protein, mitochondrial precursor.
RT SOM1 OR YEL0598C.
RT Saccharomyces cerevisiae (Baker's yeast).
RT Saccharomycetes:
RT Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
RT Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RT Nxt1_Taxid=4932;
RX SEQUENCE FROM N.A.
RA STRAIN-SC167;
RA MEDLINE=97033548; PubMed=8879245;
RA Esser K., Pratie E., Michaelis G.;
RT "SOM1, a small new gene required for mitochondrial inner membrane
RT peptidase function in Saccharomyces cerevisiae.";
RT Mol. Gen. Genet. 252:437-445(1996).
RT [1]
RX SEQUENCE FROM N.A.
RA STRAIN-S288C / AB972;
RA PubMed=9169866;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Yelton M.A., Allen E.,
RA Ardujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartwell G.,
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Laashkari D., Lew H.,
RA Lin D., Mosedale D., Nishikawa K., Namath A., Norgren R., Oeffner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RT Nature 387:78-81(1997).

```

```

CC CC -1- FUNCTION: REQUIRED FOR MITOCHONDRIAL INNER MEMBRANE PEPTIDASE
CC CC FUNCTION.
CC CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL. INNER MEMBRANE-ASSOCIATED
CC CC (POTENTIAL).
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC use by non-profit institutions only. There are no restrictions on its
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: X90459; CAA62083.1; -
DR EMBL: U18795; AAB65036.1; -
DR PIR: S72624; S72624.
DR SGD: S0002954; SOM1.
DR GO: GO:0005743; C:mitochondrial inner membrane; IDA.
DR GO: GO:0005508; P:proteolysis and peptidolysis; IGI.
DR MitoChondrion; Inner membrane; Transl peptide.
FT TRANSIT 1 74 MITOCHONDRION (POTENTIAL).
FT CHAIN 1 74 SOM1 PROTEIN.
FT SEQUENCE 74 AA; 8415 MW; BA1CBA3547954E97 CRC64;

Query Match
Best Local Similarity 25.9%; Score 59.5; DB 1; Length 74;
Matches 21; Conservative 9; Mismatches 28; Indels 23; Gaps 4;

QY 49 VRADRAVPIVNLKD-----ELFSPSEALSGSGCPKPGARIPFGCKDYLRLPT 101
DB 7 IRTDQALAPLATDSDQNCRLKELY--QWECQFKAEGAYVCSPPKRLF-----EMCI 56
QY 102 WPKSVHMGSDPNCRRLTESY 122
DB 57 APDKSA-----TWVEVTDIV 71

RESULT 4
HBB2_ANAMI STANDARD; PRT: 146 AA.
AC P83273;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin beta-2 chain.
GN HBB2.
OS Anarhichas minor (Arctic spotted wolffish).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidei;
OC Anarhichadidae; Anarhichas.
NCBI_TaxID=65739;
[1]
RP SEQUENCE, FUNCTION, AND TISSUE SPECIFICITY.
RA Pubmed-12118003;
RA Verde C., Caratore V., Riccio A., Tamburrini M., Parisi E.,
RA Di Priaso G.;
RT The functionally distinct hemoglobins of the Arctic spotted wolffish
RT Anarhichas minor.
RT J. Biol. Chem. 277:36312-36320(2002).
CC -1- FUNCTION: Involved in oxygen transport from the lung to the
CC -1- SUBUNIT: Hb3 is a heterotetramer of two alpha-2 chains and two
CC -1- beta-2 chains.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- MISCELLANEOUS: Hb3 displays pronounced Bohr and root effects,
CC -1- accompanied by strong organophosphate regulation.
CC -1- SIMILARITY: Belongs to the globin family.
DR InterPro: IPR002337; Beta.hem.
DR Pfam: PF00042; globin.1.
DR PRINTS: PR00814; BETHAEM.
DR PROSITE: PS01033; GLOBIN.1.

```

```

KW Heme; Oxygen transport; Transport; Metal-binding.
FT METAL 63 63 IRON (HEME DISTAL LIGAND) (BY
FT METAL 92 92 SIMILARITY).
FT METAL 92 92 IRON (HEME PROXIMAL LIGAND) (BY
FT SEQUENCE 146 AA; 16289 MW; 139B6730B659B800 CRC64;

Query Match
Best Local Similarity 27.5%; Score 56; DB 1; Length 146;
Matches 25; Conservative 12; Mismatches 42; Indels 12; Gaps 4;

QY 37 FUSRRLODYIVRRADRAVPIVNLKDELFPSEALSGSGCPKPGARIPFGCKGY 96
DB 6 FERATIDIFS--KMDYVGPALSRCLIVPWQRFGSGNLYNAAIIG--NPV 60
QY 97 LRHPWPQSVHMGSD---PNCGRRLTESYCE 124
DB 61 AKHGT---ILHGDGVKXNMDNIKETVAE 87

RESULT 5
Y006_DEIRA STANDARD; PRT: 159 AA.
AC Q9R9S9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein DR2406.
GN DR2406.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
NCBI_TaxID=1299;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE-20036896; PubMed-10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT *Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.
RT Science 286:1571-1577(1999).
CC -1- SIMILARITY: BELONGS TO THE UPF0152 (PAI) FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC use by non-profit institutions only. There are no restrictions on its
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: AF002071; AAF11950.1; -
DR PIR: G75277; G75277.
DR TIGR: DR2406;
DR InterPro: IPR003736; DUF157.
DR InterPro: IPR006683; Thioestr-supf.
DR Pfam: PF03061; 4HBT.1.
DR TIGRPFAM: TIGR00369; unchar_dom_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 159 AA; 17032 MW; 9EAFB596BREDBA1A CRC64;

Query Match
Best Local Similarity 21.8%; Score 55.5; DB 1; Length 159;
Matches 38; Conservative 17; Mismatches 52; Indels 67; Gaps 8;

QY 1 VALNSP-----LSGMRGIRGADPGQCGQARAVGLAGTFFAFLSSRLQDLXSYR 50

```

```
DB 7 LALPSPDEFEOLSPBALAARNNGPEGO-----GLPGLTGLARIGIR---YVSMAR 52
OY 51 RADRAAPVIVLKDLELFPSSWEALFSGSEGLPKGARIFSPD---GKDVLRHPTW-----102
DB 53 ERVAVATMV-----EGNRQPRAGRHCATLALAEELASVGSWMLND 93
OY 103 POKSVWHGSDPNG---RRLETSYCE-----TWRTAPSPATGOASSL 140
DB 94 POROVAVGVLDNGTHVRGVSSEGHVTAEARLSYRGRSLMWELIEMKDEKGRFTSL 147

RESULT 6
ACBP_YEAST STANDARD: PRT: 86 AA.
ID ACBP_YEAST STANDARD: PRT: 86 AA.
AC P31787;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
AC11-CoA-binding protein (ACBP).
ACB1 OR ACB OR YGR037C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087512; PubMed=1454809;
RA Rose T.M., Schultz E.R., Todaro G.J.;
RT "Molecular cloning of the gene for the yeast homolog (ACB) of
RT diisopropyl binding inhibitor/endopeptidase/acyl-CoA-binding protein."
RL Proc. Natl. Acad. Sci. U.S.A. 89:11287-11291(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII."
RL Yeast 13:1077-1090(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=X2180;
RX MEDLINE=96095790; PubMed=9434347;
RA Boersling C., Hummel R., Schultz E.R., Rose T.M., Pedersen M.B.,
RA Knudsen J., Kristiansen K.;
RT "Saccharomyces carlsbergensis contains two functional genes encoding
RT the acyl-CoA binding protein, one similar to the ACB1 gene from S.
RT cerevisiae and one identical to the ACB1 gene from S. monacensis."
RL Yeast 13:1409-1421(1997).
RN [4]
RP FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
RN HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
RN ACYL-COA ESTERS.
RN -1 SIMILARITY: Belongs to the ACBP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M99489; AAA34384.1; -
CC EMBL: 272822; CAAG7025.1; -
CC EMBL: Y08687; CAAG944.1; -
CC EMBL: Y08689; CAAG947.1; -
CC PIR: S31247; S31247.
CC HSSP: P07107; IACA.
CC SCD: S0003269; ACB1.
CC GO: GO:0006631; P:fatty acid metabolism; IDA.
CC InterPro: IPR000582; Ac_coa_bind_prot.
CC Pfam: PF00887; ACBP. 1.
CC PRINTS: PR00589; ACOABINDINGP.
```

```
DR ProDom: PD351532; Ac_coa_bind_prot; 1.
DR PROSITE: PS00880; ACBP. 1.
KW Transport; Lipid-binding.
FT INT MET 0 0
SQ SEQUENCE 86 AA; 9930 MW; B1BF852808EA7C3 CRC64;

Query Match 6.2%; Score 55; DB 1; Length 86;
Best Local Similarity 28.8%; Pred. No. 91;
Matches 19; Conservative 15; Mismatches 24; Indels 8; Gaps 3;

OY 20 OCFOO-ARAVGLAGFAFLSSRLDDLSIVRA-----DRAAVPIVNLKDELFPSSWEA 73
DB 3 QLEFKARAVNELPTKPS--TDELELEYALYKQATVGNDDXKGIIFMKRKYKWEANEN 60
OY 74 LFGSGE 79
DB 61 LKGKSQ 66

RESULT 7
ACB2_SACPS STANDARD: PRT: 86 AA.
ID ACB2_SACPS STANDARD: PRT: 86 AA.
AC 092272; 096496;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Acyl-CoA-binding protein 2 (ACBP type 2).
GN ACB2.
OS Saccharomyces pastorianus (Yeast), and
OS Saccharomyces monacensis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=27292; 55189;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=S.pastorianus, and S.monacensis;
RX STRAIN=AJL 2155, and CBS 1503;
RX MEDLINE=96095790; PubMed=9434347;
RA Boersling C., Hummel R., Schultz E.R., Rose T.M., Pedersen M.B.,
RA Knudsen J., Kristiansen K.;
RT "Saccharomyces carlsbergensis contains two functional genes encoding
RT the acyl-CoA binding protein, one similar to the ACB1 gene from S.
RT cerevisiae and one identical to the ACB1 gene from S. monacensis."
RL Yeast 13:1409-1421(1997).
RN [2]
RP SEQUENCE.
RX SPECIES=S.pastorianus; STRAIN=BK 2208;
RX MEDLINE=94379979; PubMed=8093000;
RA Knudsen J., Faergeman N.J., Skoelt H., Hummel R., Boersling C.,
RA Rose T.M., Andersen J.S., Hoejrup P., Roespstorff P., Kristiansen K.;
RT "Yeast acyl-CoA-binding protein: acyl-CoA-binding affinity and effect
RT on intracellular acyl-CoA pool size."
RL Biochem. J. 302:479-485(1994).
RN [3]
RP FUNCTION: BINDS MEDIUM- AND LONG-CHAIN ACYL-COA ESTERS WITH VERY
RN HIGH AFFINITY AND MAY FUNCTION AS AN INTRACELLULAR CARRIER OF
RN ACYL-COA ESTERS (BY SIMILARITY).
RN -1 SIMILARITY: Belongs to the ACBP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y08690; CAAG9948.1; -
CC EMBL: Y08688; CAAG9946.1; -
CC HSSP: P07107; IACA.
CC InterPro: IPR000582; Ac_coa_bind_prot.
CC Pfam: PF00887; ACBP. 1.
CC PRINTS: PR00589; ACOABINDINGP.
CC ProDom: PD351532; Ac_coa_bind_prot; 1.
```

DR PROSITE: PS00880; ACP; 1.
 KW Transport; Lipid-binding.
 FT INIT_MER 0
 SQ SEQUENCE 86 AA; 9953 MW; 72B9989A4EC7C21 CRC64;

Query Match
 Best Local Similarity 28.8%; Score 54; DB 1; Length 86;
 Matches 19; Conservative 14; Mismatches 25; Indels 8; Gaps 3;

OY 20 QCFQO-ARAVGLAGTFRFLSLRDLYSIVRA-----DRAAVIVINIKDELLPSPWEA 73
 DB 3 QLFEEKAKAVNELPKPS--IDELLELYGLYKQATVGNDKEKPGIFNMKDRYKWEAWED 60
 OY 74 LPSGSE 79
 DB 61 LKGSQ 66

RESULT 8

MERC_THIPE STANDARD: PRT; 144 AA.
 P22905;

DT 01-AUG-1991 (Rel. 19, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Mercutic resistance protein merc.
 GN MERC.
 OS Thiobacillus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 OX NCBI_TaxID=920;

RA STRAIN-E-15;
 RX MEDLINE=91092487; PubMed=2265748;
 RA Inoue C., Sugawara K., Kusano T.;
 RT "Thiobacillus ferrooxidans mer operon: sequence analysis of the
 RT promoter and adjacent genes.";
 RL Gene 96:115-120(1990).
 RN [2]

RP REVISION TO 130.
 RA Inoue C., Sugawara K., Kusano T.;
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN IS LOCATED IN THE INNER MEMBRANE AND
 CC MEDIATES HG(2+) TRANSPORT INTO THE CYTOPLASM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- INDUCTION: BY HG(2+).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D90110; BAA14138.2; -
 DR PIR; JS0479; JS0479.
 DR InterPro: IPR004891; Merc.
 DR Pfam: PF03203; Merc; 1.
 DR Transport; Transmembrane; Inner membrane; Mercury;
 KW Mercutic resistance.
 FT TRANSMEM 11
 FT TRANSMEM 47
 FT TRANSMEM 67
 FT TRANSMEM 79
 FT TRANSMEM 99
 FT TRANSMEM 107
 FT TRANSMEM 127
 FT TRANSMEM 144
 FT SEQUENCE 144 AA; 15526 MW; CD6829E430FFB898 CRC64;

Query Match
 Best Local Similarity 25.7%; Score 54; DB 1; Length 144;
 Matches 18; Conservative 7; Mismatches 25; Indels 20; Gaps 2;

OY 71 WEALFSGSGEPL-----KPGARIFSFQKDVLRHPTWPKSVWHSQDPNG 115

DB 75 WORTLLGSLGPVLAIVGVGLTHHFLDKLARIVFTGLVWF-----LVSTMDVYNPAN 129
 OY 116 RLRESYCEPT 125
 DB 130 RRCATDCGET 139

RESULT 9

RK16_CYAPA STANDARD: PRT; 136 AA.
 ID RK16_CYAPA
 AC P23406;

DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cyanelle 50S ribosomal protein L16.
 GN RPL16.
 OS Cyanophora paradoxa.
 OC Cyanelle.
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
 OX NCBI_TaxID=2762;

RA STRAIN-UTEX LB 555 / Pringsheim;
 RX MEDLINE=91117189; PubMed=2126059;
 RA Michalowski C.B., Pfanzagl B., Loeffelhardt W., Bohnert H.J.;
 RT "The cyanelle S10 spc ribosomal protein gene operon from Cyanophora
 RT paradoxa.";
 RL Mol. Gen. Genet. 224:222-231(1990).
 RN [2]

RP SEQUENCE FROM N.A.
 RA STRAIN-UTEX LB 555 / Pringsheim;
 RA Strewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
 RA Bryant D.A.;
 RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
 RL Plant Mol. Biol. Rep. 13:327-332(1995).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-UTEX LB 555 / Pringsheim;
 RC Loeffelhardt W., Strewalt V.L., Michalowski C.B., Annarella M.,
 RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
 RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
 RT the genetic complexity of a primitive plastid.";
 RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
 RL Schwenmler W. (eds.);
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
 RL (1997).

CC -1- SIMILARITY: BELONGS TO THE L16P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M30487; AAA63622.1; -
 DR EMBL; U30821; AAA81226.1; -
 DR PIR; S1213; RSKT16.
 DR InterPro: IPR000114; Ribosomal_L16.
 DR Pfam: PF00252; Ribosomal_L16; 1.
 DR PRINTS: PR00060; RIBOSOMALL16.
 DR TIGRFS: TRG01164; rpl16; 1.
 DR PROSITE: PS00586; RIBOSOMAL_L16; 1.
 DR PROSITE: PS00701; RIBOSOMAL_L16_2; 1.
 KW Ribosomal protein; Cyanelle.
 SQ SEQUENCE 136 AA; 15537 MW; 9E07D199C9F50DD4 CRC64;

Query Match
 Best Local Similarity 27.2%; Score 53.5; DB 1; Length 136;
 Matches 28; Conservative 13; Mismatches 35; Indels 27; Gaps 5;

OY 9 GGMRCI--RG-----ADFOCEQOARAVGLAGTFFRAFLSSR-----LODLYSIVRADRAA 56
 DB 15 GRMGISIRGNLVGDF-----GLQALEPAMISROLEASRRALNRYVRGKIM 65
 OY 57 VPIVNLKDELLEPPSWALEPFSGSEGP-----LKPGARIFPSGDC 93
 DB 66 IRIFPDKPVTMRPAETRMGSGKGAPEYWAIVKPGRVIFEING 108

RESULT 10

MUCA_SALTY STANDARD: PRT: 146 AA.
 AC P07376; P14302;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE MUCA protein (EC 3.4.21.-) [Contains: MUCA' protein].
 GN MUCA.
 SS Salmoneella typhimurium, and
 SS Escherichia coli.
 OG Plasmid Incn PKM101, and Plasmid Incn R46.
 OC Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OX Enterobacteriaceae; Salmoneella.
 RN NCB1_TaxID=602, 562;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC PLASMID-INCEN PKM101;
 RA Tanooka H.;
 RL Submitted (MAY-1990) to the PIR data bank.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC PLASMID-INCEN PKM101;
 RA MEDLINE=85242678; PubMed=2989816;
 RA Perry R.L., Ellledge S.J., Mitchell B.B., Marsh L., Walker G.C.;
 RT "uninduced and mucab operons whose products are required for UV light-
 and chemical-induced mutagenesis: umud, mucA, and LexA proteins share
 RT homology.";
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4331-4335(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC PLASMID-INCEN R46;
 RA Hall R.M., Vockler C.;
 RL Submitted (SEP-1989) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN UV PROTECTION AND MUTATION.
 CC -1- MISCELLANEOUS: THE MUCAB OPERON IS THE PLASMID-BORNE ANALOG OF THE
 CC E. COLI UMUDC OPERON.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S24.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D90147; BAAL1475.1; -;
 CC EMBL: X16596; CAA34606.1; -;
 CC EMBL: M13388; AAA98277.1; -;
 CC PIR: D23157; ZMECAP.
 CC PIR: S06775; S06775.
 CC HSSP: P04153; IAY9.
 CC MEROPS: S24.003; -;
 CC InterPro: IPR006197; Peptidase_S24.
 CC InterPro: IPR006198; Peptidase_S24_C.
 CC Pfam: PF00717; Peptidase_S24; 1.
 CC PRINTS: PR00726; LEXASERPTASE.
 CC Plasmid: SOS mutagenesis; DNA repair; Autocatalytic cleavage;
 CC Hydrolase; Serine protease.
 CC PROPEP 1 26 BY SIMILARITY.
 FT CHAIN 27 146 MUCA' PROTEIN.
 FT SITE 26 27 CLEAVAGE (AUTO-).

FT ACT_SITE 62 62 INVOLVED IN AUTO-CLEAVAGE
 FT ACT_SITE 99 99 (BY SIMILARITY).
 FT CONFLICT 13 13 INVOLVED IN AUTO-CLEAVAGE
 FT CONFLICT 13 13 (BY SIMILARITY).
 SQ SEQUENCE 146 AA; 16527 MW; 7605A681F8345E76 CRC64;
 Query Match 6.0%; Score 53.5; DB 1; Length 146;
 Best Local Similarity 30.4%; Pred. No. 2.4e+02;
 Matches 17; Conservative 7; Mismatches 17; Indels 15; Gaps 3;

OY 122 YCEWTWTEAPSAT-----GQASSLGGRL-----LGQASASCHHAYVLCTENSF.166
 DB 45 YC-----VHNPATATYLRVSGSMEDGRIDGVDLVVDRSLTASHGSIYVACIHNEF 96

RESULT 11

C354_RHOSH STANDARD: PRT: 153 AA.
 ID C554_RHOSH
 AC 053142;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytochrome c-554 precursor (C354) (High potential cytochrome c).
 GN CYC.
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria: Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OX Rhodobacteraceae; Rhodobacter.
 RN NCB1_TaxID=1063;
 RP [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
 RA MEDLINE=95362655; PubMed=7543472;
 RA Flory J.E., Donohue T.J.;
 RT "Organization and expression of the Rhodobacter sphaeroides cycC
 RT operon.";
 RT J. Bacteriol. 177:4311-4320(1995).
 RN [2]
 RP SEQUENCE OF 21-153.
 RC STRAIN-MRE;
 RX MEDLINE=89271926; PubMed=2543295;
 RA Bartsch R.G., Ambler R.P., Meyer T.E., Cusanovich M.A.;
 RT "Effect of aerobic growth conditions on the soluble cytochrome
 RT content of the purple phototrophic bacterium Rhodobacter sphaeroides:
 RT induction of cytochrome c554.";
 RL Arch. Biochem. Biophys. 271:433-440(1989).
 CC -1- FUNCTION: MONOHEME C-TYPE CYTOCHROME, THAT IS PARTICULARLY
 CC EXPRESSED WHEN CELLS GENERATE ENERGY VIA AEROBIC RESPIRATION.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- PTM: BINDS ONE HEME GROUP PER MOLECULE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: L36880; AAD09145.1; -;
 CC HSSP: P00149; IATV.
 CC InterPro: IPR002321; CyC_CII.
 CC InterPro: IPR000345; CyC_heme_bind.
 CC Pfam: PF01322; Cytochrome_C_2; 1.
 CC PRINTS: PR00608; CYTCROME_CII.
 CC PRODOM: PD003828; CyC_CII; 1.
 CC PRINTS: PS00190; CYTOCHROME_C; 1.
 CC Electon transport; Heme; Periplasmic; signal;
 CC Pyrolydnone carboxylic acid.
 CC SIGNAL 1 20
 FT CHAIN 21 153 CYTOCHROME C-554.
 FT MOD_RES 21 21 PYROLYDNONE CARBOXYLIC ACID (PROBABLE).
 FT BINDING 142 142 HEME (COVALENT) (BY SIMILARITY).

FT BINDING 145 145 HEME (COVALENT) (BY SIMILARITY).
 FT METAL 37 37 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT METAL 146 146 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT CONFLICT 42 42 H -> Y (IN REF. 2).
 FT CONFLICT 64 64 K -> D (IN REF. 2).
 FT CONFLICT 76 76 I -> L (IN REF. 2).
 FT CONFLICT 112 112 G -> D (IN REF. 2).
 FT CONFLICT 141 141 A -> G (IN REF. 2).
 SQ SEQUENCE 153 AA; 16073 MW; 421EB28F08C42200 CRC64;

Query Match 6.0%; Score 53.5; DB 1; Length 153;
 Best Local Similarity 19.8%; Pred. No. 2.5e+02;
 Matches 18; Conservative 12; Mismatches 40; Indels 21; Gaps 1;

OY 67 LEPSWEALFSGSEGPLKPGARIFSPDKDVLHPTWPKSVHSGSDPNGRRLTESYCETWTEA 126
 DB 80 LEFGSAVGEAEDSALP-----ALWEKPDDEFRQKISGMEEA 118
 OY 127 RTEAPSATGQASSLLGRLLGQSAASCHHAY 157
 119 KMOAAGTDLASLQCAMRDLAGACSGCHETY 149

RESULT 12

CYCP_RHOSP STANDARD: PRT: 128 AA.
 AC P00146;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-MAR-1989 (Rel. 10, Last annotation update)
 DE Cytochrome c'.
 OS Rhodopseudomonas sp.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiales; Rhodopseudomonas.
 ON NCBI_TaxID=1078;
 RX MEDLINE=82082545; PubMed=6273892;
 RA Ambler R.P., Bartsch R.G., Daniel M., Kamen M.D., Mclellan L.,
 RT Meyer T.E., van Beeumen J.,
 RL "Amino acid sequences of bacterial cytochromes c' and c-556.";
 CC Proc. Natl. Acad. Sci. U.S.A. 78:6854-6857(1981).
 CC -1- FUNCTION: CYTOCHROME C' IS THE MOST WIDELY OCCURRING BACTERIAL
 C-TYPE CYTOCHROME. CYTOCHROMES C' ARE HIGH-SPIN PROTEINS AND THE
 HEME HAS NO SIXTH LIGAND. THEIR EXACT FUNCTION IS NOT KNOWN.
 DR HSP: P00147; CCRFCX.
 DR InterPro: IPR002321; Cyt_CII.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR Pfam: PF01322; Cytochrome_c_2; 1.
 DR PROSITE: PS003828; Cyt_CII; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 FT BINDING 117 117 HEME (COVALENT).
 FT BINDING 120 120 HEME (COVALENT).
 FT METAL 121 121 IRON (HEME AXIAL LIGAND).
 SQ SEQUENCE 128 AA; 13371 MW; F49957285708124F CRC64;

Query Match 5.9%; Score 53; DB 1; Length 128;
 Best Local Similarity 25.8%; Pred. No. 2.3e+02;
 Matches 23; Conservative 13; Mismatches 43; Indels 10; Gaps 4;

OY 72 EALFSGSEGPL-KPGARIFSPDKDVLHPTWPKSVHSGSDPNGRRLTESYCETWTEA 130
 DB 43 EALILATDIFLPLRPGTSDADPESEPAK-----ASIWNMEDFGAKGQAMHEAGWELIA 96
 OY 131 PSATGQASSLLGRLLGQSAASCHHAY 157
 DB 97 AANTGEASA-FGPALKKIGGTCKACHDDY 124

RESULT 13

HBAD COLLI
 ID HBAD_COLLI STANDARD: PRT: 140 AA.
 AC 012985;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hemoglobin alpha-D chain.

OS Columba livia (Domestic pigeon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
 ON NCBI_TaxID=8932;
 RX MEDLINE=97320440; PubMed=9177291;
 RA Ikehara T., Eguchi Y., Kayo S., Takel H.;

RT "Isolation and sequencing of two alpha-globin genes alpha(A) and
 alpha(D) in pigeon and evidence for embryo-specific expression of the
 alpha(D)-globin gene.";

RL Biochem. Biophys. Res. Commun. 234:450-453(1997).
 CC -1- FUNCTION: Involved in oxygen transport from the lung to the
 various peripheral tissues.

CC -1- SUBUNIT: Heterotrimer of two alpha-D chains and two beta chains.
 CC -1- TISSUE SPECIFICITY: Red blood cells.

CC -1- DEVELOPMENTAL STAGE: In birds, the alpha-D chain occurs in a minor
 late embryonic and adult life.

CC -1- SIMILARITY: Belongs to the globin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AB001981; BAA19668.1; .
 DR PIR: JC5515; JC5515.
 DR HSP: P02001; IHRB.
 DR InterPro: IPR002338; Alpha_haem.

DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin; 1.
 DR PRINTS: PR00612; ALPHAHAEEM.
 DR PROSITE: PS01033; GLOBIN; 1.

DR Heme; Oxygen transport; Transport; Erythrocyte.
 FT METAL 57 57 IRON (HEME DISTAL LIGAND).
 FT METAL 86 86 IRON (HEME PROXIMAL LIGAND).
 SQ SEQUENCE 140 AA; 15717 MW; 2FAC884799A152P9 CRC64;

Query Match 5.9%; Score 53; DB 1; Length 140;
 Best Local Similarity 26.2%; Pred. No. 2.5e+02;
 Matches 22; Conservative 8; Mismatches 20; Indels 34; Gaps 4;

OY 96 VLRHP-----TWPKSVW-----HGSDP-----NGRRLTESYCETWTEAPSA 133
 DB 17 VLRHPGCAEALRLFTYRQTKYFRRPHDLHGSDQVRRHNGKVL-----AA 64
 OY 134 TGOASSLLGRLLGQSAASCHHAY 157
 DB 65 LGNAVKSLGNLSQALSDLDLHAY 88

RESULT 14

UBCJ HUMAN
 ID UBCJ_HUMAN STANDARD: PRT: 165 AA.
 AC P56354;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin-conjugating enzyme E2 G2 (Ec 6.3.2.19) (ubiquitin-protein
 DE ligase G2) (ubiquitin carrier protein G2).
 GN UBE2G2.

OS Homo sapiens (Human), and
 OC Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606, 10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human;
 RX MEDLINE=98360100; PubMed=9693041;
 RA Katsanis N., Fisher E.M.C.;
 RT "Identification, expression, and chromosomal localization of ubiquitin
 conjugating enzyme 7 (UBE2C2), a human homologue of the Saccharomyces
 cerevisiae UBC7 gene.";
 RL Genomics 51:128-131(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human;
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hatori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 Park H.-S., Toyoda A., Ishii K., Tokoi Y., Choi D.-K., Groner Y.,
 Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 Reichwald K., Rump A., Schilhabel M., Schudy A., Zimmermann W.,
 Rosenthal A., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,
 Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 Ransner J., Beck A., Klages S., Hennig S., Rieselmann L., Dgand E.,
 Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 Lehman H., Reinhardt R., Taspo M.-L.;
 RA "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Human; TISSUE=Lung;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Mouse; STRAIN=C57BL/6J; TISSUE=Fetal;
 RX MEDLINE=21238294; PubMed=11278356;
 RA Tiwari S., Weissman A.M.;
 RT "Endoplasmic reticulum (ER)-associated degradation of T cell receptor
 subunits. Involvement of ER-associated ubiquitin-conjugating enzymes
 (E2s).";
 RL J. Biol. Chem. 276:16193-16200(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=Mouse;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO
 OTHER PROTEINS.
 CC -1- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine - AMP +
 diphosphate + protein N-ubiquityllysine.
 CC -1- PATHWAY: Ubiquitin conjugation; second step.
 CC -1- MISCELLANEOUS: A cysteine residue is required for ubiquitin-
 thiolester formation.
 CC -1- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.
 CC STRONGEST: NO C. ELEGANS UBC7.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF032456; AAC32312.1; -
 DR EMBL: AL163300; CAB90551.1; -
 DR EMBL: BC001738; AAH01738.1; -
 DR EMBL: BC008351; AAH08351.1; -

DR EMBL: BC011569; AAH11569.1; -
 DR EMBL: AF296657; AAH52608.1; -
 DR EMBL: BC010321; AAH10321.1; -
 DR HSSP: 002159; 20C2.
 DR Genew: HGNC:12483; UBE2C2.
 DR MIM: 603124; -
 DR MGD: MGI:1343188; ube2g2.
 DR GO: GO:0004840; F:ubiquitin conjugating enzyme activity; TAS.
 DR InterPro: IPR000608; UBC_conjugat.
 DR Pfam: PF00179; UQ_con; 1.
 DR Prodom: PD000461; UBC_conjugat; 1.
 DR SMART: SM00212; UBCc; 1.
 DR PROSITE: PS00183; UBIQUITIN_CONUGAT_1; 1.
 DR PROSITE: PS0127; UBIQUITIN_CONUGAT_2; 1.
 DR UBI conjugation pathway; Ligase; Multigene family.
 FT BINDING 89 UBIQUITIN (BY SIMILARITY).
 FT CONFLICT 12 12 E -> V (IN REF. 1).
 FT CONFLICT 101 107 MGYESA -> HGLREOP (IN REF. 1).
 SQ SEQUENCE 165 AA: 18566 MW: 74DEC732A79575E3 CRC64;
 Query Match 5.9%; Score 53; DB 1; Length 165;
 Best Local Similarity 28.6%; Pred. No. 3e+02;
 Matches 24; Conservative 5; Mismatches 23; Indels 32; Gaps 6;
 QY 68 PPSWEALFSGSEG-----PLKFGARIFSGDKVLRHPT-WPQK--- 105
 DB 34 FEEMALIMGPEDFCEFGVFPALIFPLDPLSPKMRFCFCE---MFHNIYDPGRVC 89
 QY 106 -SVNH-GSDPRNGRLTFESYCETW 126
 DB 90 ISILHAPGDDPWG---YESSAERW 110
 RESULT 15
 UBC7_SCHPO STANDARD: PRT; 166 AA.
 AC 000102; Q9HDP3; STRAIN=972;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin-conjugating enzyme E2-18 kDa (EC 6.3.2.19) (ubiquitin-
 protein ligase) (ubiquitin carrier protein).
 GN UBCP3 OR SPBP16F5.04.
 OS Schizosaccharomyces pombe (fission yeast).
 CC Schizosaccharomycetes; Schizosaccharomycetaceae;
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetaceae;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 OC NCBI_TaxID=4896;
 OX [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=9729689; PubMed=9154838;
 RA Osaka F., Setno H., Seno T., Yamao F.;
 RT "A ubiquitin-conjugating enzyme in fission yeast that is essential
 for the onset of anaphase.";
 RL Mol. Cell. Biol. 17:3388-3397(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Sgouros K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voickert G., Aert R., Robben J., Glynnoprez B.,



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:16:23 ; Search time 94 Seconds

(Without alignments)
466.691 Million cell updates/sec

Title: US-09-171-607A-1

Perfect score: 893
Sequence: 1 VALNSPLSGNGRGIRGADFO.....ASCHHAYIVLCIENSFWMTAS 170

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 295467

Minimum DB seq length: 0
Maximum DB seq length: 170

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeop:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	722	80.9	160	11	09CRT2 mus musculus
2	529	59.2	102	4	096RT70 homo sapien
3	70.5	7.9	91	7	019588 homo sapien
4	70.5	7.9	165	7	09MW45 homo sapien
5	69.5	7.8	168	2	045813 mycobacteri
6	68.5	7.7	154	4	09NM01 homo sapien
7	67	7.5	147	13	09DGR8 xenopus lae
8	65.5	7.3	108	3	09HPS4 coccidioid
9	65	7.3	149	5	09NED6 leishmania
10	64	7.2	126	17	026681 methanobact
11	64	7.2	170	4	08RT2 rhizobium m
12	63	7.1	113	16	092257 pan troglod
13	62.5	7.0	78	7	031004 dirosophila
14	62.5	7.0	116	5	09YAR4 homo sapien
15	62.5	7.0	127	4	08N860 arabadopis
16	62.5	7.0	146	10	09SXS4 arabadopis

17	62	6.9	102	4	08WY95
18	62	6.9	126	17	08RT72
19	62	6.9	137	10	08SB91
20	62	6.9	147	5	061579
21	62	6.9	158	3	000171
22	61	6.8	78	6	08S084
23	61	6.8	167	7	046771
24	60.5	6.8	95	9	08SDG4
25	60.5	6.8	118	11	08C567
26	60.5	6.8	154	4	09NRY2
27	60.5	6.8	156	16	09A9N5
28	60.5	6.8	161	5	08IS35
29	60	6.7	131	4	09ENCO
30	60	6.7	135	2	09L884
31	60	6.7	151	17	09HPE1
32	60	6.7	152	17	09YDH5
33	60	6.7	157	9	09T140
34	59.5	6.7	84	16	08XS09
35	59.5	6.7	87	10	08S2C8
36	59.5	6.7	146	12	08JU56
37	59	6.6	83	16	09K6W7
38	59	6.6	125	12	0919J3
39	59	6.6	135	16	08KCG0
40	59	6.6	150	16	08YXB5
41	59	6.6	160	11	09DA24
42	59	6.6	160	17	08TID9
43	58.5	6.6	167	5	09VXE8
44	58.5	6.6	83	9	037933
45	58.5	6.6	90	2	09AE38

ALIGNMENTS

RESULT 1

09CRT2 PRELIMINARY; PRT; 160 AA.

ID 09CRT2

AC 09CRT2; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DR Procollagen, type XVIII, alpha 1 (Fragment).

GN COL18A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RC SEQUENCE FROM N.A.

RX STRAIN=C57BL/6J; TISSUE=Embryo;

MDLINE=21085660; Pubmed=11217851;

RA Kawal J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., Kling B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T., Blake J., Boileau D., Hojunga N., Carninci P., de Bonaldo M.F., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.,

Functional annotation of a full-length mouse cDNA collection.;

RT Nature 409:685-690(2001).

RL EMBL: AK014292; BAB29249.1; -.

DR HSP; P39061; IKOE.

DR MGI:88451; Coll18a1.
RT NON_TER 1
SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;

Query Match
Best Local Similarity 80.9%; Score 722; DB 11; Length 160;
Matches 134; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

OY 13 GIRGADFCFOQARAVGLAGTFRAFLSRLDLYSIYRADAIVYNLKLDPSPWE 72
DB 1 GIRGADFCFOQARAVGLSGTFRAFLSRLDLYSIYRADAIVYNLKLDPSPWD 60
OY 73 ALFSGSEGLPKPGARISFDKDVLRHPTWPKSVHMGSDPNGRRLTESYCEWTAPAS 132
DB 61 SLFSGSQQLQPGARISFDGRDVLRRHPWPKSVHMGSDPNGRRLMESYCEWTETGTG 120
OY 133 ATGQASSILGRLGQASASCHAHAYIVLCIENSRMTA 169
DB 121 ATGQASSILGRLLEQKAAKASCHNSYIVLCIENSRMTS 157

RESULT 2

O96T70 PRELIMINARY; PRT; 102 AA.

AC O96T70;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Endostatin variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Deininger M.H., Trautmann K., Schluesener H.J.;
RT "Endostatin promotes delayed secondary damage following traumatic brain injury."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF33247; AAK50626.1; -
FT NON_TER 1
FT SEQUENCE 102 AA; 11147 MW; ECAC47AA6420947D CRC64;

Query Match
Best Local Similarity 98.0%; Score 529; DB 4; Length 102;
Matches 100; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 59 IVNLKDELPPSWALFSGSEGLPKPGARISFDKDVLRHPTWPKSVHMGSDPNGRRL 118
DB 1 IVNLKDELPPSWALFSGSEGLPKPGARISFDKDVLRHPTWPKSVHMGSDPNGRRL 60
OY 119 TESYCEWTAPASATGQASSILGRLGQASASCHAHAYIVL 160
DB 61 TESYCEWTAPASATGQASSILGRLGQASASCHAHAYIVL 102

RESULT 3

ID O19588 PRELIMINARY; PRT; 91 AA.

AC O19588;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MHC class I antigen (Fragment).
CN HLA-G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=98041190; PubMed=9373856;

RA Hvald T.V., Meldgaard M., Sorensen S., Morling N.;
RT "Polymorphism of exon 3 of the HLA-G gene."
RL J. Reprod. Immunol. 35:31-42(1997).
DR EMBL; AF020714; AAC51863.1; -
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00129; MHC_I.1.
DR ProDom; PD000050; MHC_I; 1.
FT NON_TER 1
FT SEQUENCE 91 AA; 10422 MW; CASC0D9A91B39F29 CRC64;

Query Match
Best Local Similarity 7.9%; Score 70.5; DB 7; Length 91;
Matches 21; Conservative 5; Mismatches 15; Indels 41; Gaps 3;

OY 77 GSEGLPKGARIISFDKDVLRHPTWPKSVHMGSDPNGRRLTESYCEWTETGTG 106
DB 13 GSDGRLRGYDQYAVDGKDYALNEDCAPGPORTLRRLSPASVRRPMMLNKGEPTWRAR 72
OY 107 VMHGSDPNGRRLTESYCEWTETGTG 128
DB 73 AMSGS-----TDWTWT 83

RESULT 4

O9MM45 PRELIMINARY; PRT; 165 AA.

AC O9MM45;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE HLA-G cell surface glycoprotein (Fragment).
CN MHC-G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Arnalitz-Villena A., Martinez-Laso J., Alvarez M., Castro M.J.,
RA Varela P., Gomez-Casado E., Suarez B., Recio M.J., Vargas-Alarcon G.,
RA Morales P.;
RT "Primate Mhc-E and -G alleles."
RL Immunogenetics 46:251-266(1997).
RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE-Blood;
RX MEDLINE=97246746; PubMed=9089111;
RX Suarez M.B., Morales P., Castro M.J., Fernandez V., Varela P.,
RA Alvarez M., Martinez-Laso J., Arnalitz-Villena A.;
RT "A new HLA-G allele (HLA-G*0105N) and its distribution in the Spanish population."
RL Immunogenetics 45:464-465(1997).
DR EMBL; L78073; AAF25961.1; -
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00129; MHC_I.1.
DR PRINTS; PR01638; MHCCLASSI.
DR ProDom; PD000050; MHC_I; 1.
FT NON_TER 1
FT SEQUENCE 165 AA; 18986 MW; 153474B1A7998A19 CRC64;

Query Match
Best Local Similarity 7.9%; Score 70.5; DB 7; Length 165;
Matches 21; Conservative 5; Mismatches 15; Indels 41; Gaps 3;

OY 77 GSEGLPKGARIISFDKDVLRHPTWPKSVHMGSDPNGRRLTESYCEWTETGTG 106
DB 94 GSDGRLRGYDQYAVDGKDYALNEDCAPGPORTLRRLSPASVRRPMMLNKGEPTWRAR 153
OY 107 VMHGSDPNGRRLTESYCEWTETGTG 128

Db 154 AMWS-----TDTWRT 164

RESULT 5

ID 049813 PRELIMINARY; PRT; 168 AA.

AC 049813: 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE B2168_C2_214.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith D.R.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U00018: AAA1724.1;
 SQ SEQUENCE 168 AA; 18757 MW; BA37547945A6A7F6 CRC64;

Query Match 7.8%; Score 69.5; DB 2; Length 168;
 Best Local Similarity 29.5%; Pred. No. 36;
 Matches 36; Conservative 11; Mismatches 44; Indels 31; Gaps 6;

Db 3 LNSPLSGMGRIGADFCQOARAVGL-----AGTRAFLSRLQDLYSIVRAD 53
 14 LAGLVLPTMCG-RDARQLLHEODALGCAAMTKAGCTIN-IGASIMMLSOAIRAG 71
 QY 54 RAAVPI-----VNLKDELFPSSWEALFSGSEGLPKGARIFFSGDKVLRNP 100
 72 RIAPFIRSGVWALDLRWVNRVETIRDOFEYL---SYGVMVTTRM---GSELGYH 124
 QY 101 TW 102
 Db 125 KW 126

RESULT 6

ID 09NM01 PRELIMINARY; PRT; 154 AA.

AC 09NM01: 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Hypothetical protein FLJ20694.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-ileal mucosa;
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK000701: BAA91328.1;
 KW Hypothetical protein.
 SQ SEQUENCE 154 AA; 16981 MW; FEC376A96991C823 CRC64;

Query Match 7.7%; Score 68.5; DB 4; Length 154;
 Best Local Similarity 27.7%; Pred. No. 41;
 Matches 23; Conservative 7; Mismatches 36; Indels 17; Gaps 3;

QY 77 GSEGLPKGARIFFS-----DGKDVLRHPTWPKSVWGH-----SDPNCRRLTESYC 123

Db 2 GRSPFKRNKVGFSYPCWCRSYPPFPRKRAMPPSRVWLGACCASLASPPKGTIPSGEY 61
 QY 124 ETRTEAPSATGQASSLGRLL 146
 Db 62 R---PAPSSGDSLRESGALL 80

RESULT 7

ID 09DGK8 PRELIMINARY; PRT; 147 AA.

AC 09DGK8: 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Atrial natriuretic factor.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Small E.M., Krieg P.A.;
 RT "Expression of atrial natriuretic factor (ANF) during Xenopus cardiac
 development."
 RT development.
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF287050: AAG01000.1;
 DR InterPro: IPR006653: Natri-peptide.
 DR Pfam: PF00212: ANP; 1.
 DR PRINTS: PR00710; Natripeptides.
 DR SMART: SM00183; NAT_PEP. 1.
 DR PROSITE: PS00263; NATRIURETIC_PEPTIDE. 1.
 SQ SEQUENCE 147 AA; 16350 MW; 99B6C26778830FD CRC64;

Query Match 7.5%; Score 67; DB 13; Length 147;
 Best Local Similarity 21.8%; Pred. No. 55;
 Matches 31; Conservative 25; Mismatches 40; Indels 46; Gaps 8;

Db 34 FRAFLSRLQDLYSIVRRADRAVPIVNLKDELFPSSWEALFSGSEGLPKGARIFF--SF 91
 27 YSSVLSDDLDDKILNLER-----LEDRV-----AEEPMAPSQDLFAQNY 66
 QY 92 DGKDVLRH-PTWPKSVWGHSDPNGRRLTESYCEFWRT-----EAPSATG 135
 67 DAADSSNAPSMTGEAIRPSD-----IIVKSSWETPKLSRLKSLRELLNSPRLR 121
 Db 122 RSSDCEFGRIDRIGAGSGMGCN 143

RESULT 8

ID 09HFS4 PRELIMINARY; PRT; 108 AA.

AC 09HFS4: 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase)
 (fragment).
 OS Coccidioides immitis.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Onygenales; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Onygenales; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 NCBI_TaxID=5501;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delgado N., Cole G.T.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) -> ADP + PHOSPHATE +
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

```

DR EMBL: AF306534; AAG28781.1; -
DR InterPro: IPR001691; GLN_synth.
DR Pfam: PF03951; gln-synt_N; 1.
DR PROSITE: PS00180; GLNA_1; 1.
KW Ligase.
FT NON_TER
SQ SEQUENCE 108 AA; 12122 MW; EEC04AF4C3DB52B3 CRC64;

Query Match
Best Local Similarity 30.0%; Score 65.5; DB 3; Length 108;
Matches 27; Conservative 6; Mismatches 34; Indels 23; Gaps 7;

OY 44 DLYSTRADRAAVIYNLKDLELPSSWEALFSGSEGLPKPGARLFSFGKDYLRHT-- 101
DB 33 DAYGVSRTKTKVNVKVSADLE--PEWN--FDGSSSTGOAPG-----DMSDYLRPAVI 82
OY 102 WPKSVHMGSDP--NGRRLTSEYCEFWRE 129
DB 83 FP-----DPRRGENTL-VLCETWSD 103

RESULT 9
O9NED6 PRELIMINARY; PRT; 149 AA.
AC O9NED6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 15.6 kDa protein.
GN P1105.07
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxId=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Frieldlin;
RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
RA Rajandream M.A., Barrell B.G.;
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Frieldlin;
RA MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagheradeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RA "A physical map of the Leishmania major Frieldlin genome.";
RT Genome Res. 8:135-145(1998).
DR EMBL: AL139794; CAB75565.1; -
DR InterPro: IPR007017; DUF657.
DR Pfam: PF04933; DUF657; 1.
DR Hypothetical protein.
SQ SEQUENCE 149 AA; 15563 MW; 647A6E3D95B6B9F0 CRC64;

Query Match
Best Local Similarity 30.1%; Score 65; DB 5; Length 149;
Matches 25; Conservative 8; Mismatches 32; Indels 18; Gaps 4;

OY 77 GSEGPLKPGARLFSFGKDYLRHTWPKSVHMGSDPNRRLTSEYCEFWREAPSATGO 136
DB 7 GGGGGLRGSGRGSGSFGFAERRP-----CTSHRPQPNQ--VTNLY-----VGRPMAGG 55
OY 137 ASSLL-----GRRLLGOSAS 152
DB 56 GSGIIGTMAAVAGGSVLGHGHSN 78

RESULT 10
O26681 PRELIMINARY; PRT; 126 AA.
AC O26681;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

```

```

DE Hypothetical protein MTH581.
GN MTH581.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxId=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadator A., Viscer R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bush D., Sater H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000840; AAB85087.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 126 AA; 14243 MW; 629A20B6C4BA0E33 CRC64;

Query Match
Best Local Similarity 32.3%; Score 64; DB 17; Length 126;
Matches 21; Conservative 10; Mismatches 30; Indels 4; Gaps 3;

OY 40 SRLQDYSIVRRADRAAVIYNLKDLELP--PSWEALFSGSEGLPKPGARLFSF-DGKD-- 95
DB 17 AELSDIRLEFRKANLHGACVKKLTIDNISKEHEANFRMSDVRHAAVFALSDGGDLE 76
OY 96 VLRRP 100
DB 77 VLXEP 81

RESULT 11
O8TER2 PRELIMINARY; PRT; 170 AA.
AC O8TER2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE FLJ00091 protein (Fragment).
GN FLJ00091.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK074039; BAB84865.1; -
FT NON_TER
SQ SEQUENCE 170 AA; 17827 MW; C8EBEAFB3ADP236 CRC64;

Query Match
Best Local Similarity 27.5%; Score 64; DB 4; Length 170;
Matches 28; Conservative 6; Mismatches 40; Indels 28; Gaps 7;

OY 84 PGARLFSFGKD--VLRIPTWPKSVY----HGSPPNRRLTSEYCE-----ETW 126
DB 72 PGGVSTSDGGHRAVLNHRPLW--VPGVWLHLRDEGQGRP---WCWGDHAGCVAAPGTW 126
OY 127 RTE----APSATGQASSLLG--GRLLGOSASCHAAVYVLCI 162
DB 127 HPAMPSPSPGLGNPQHLNHRGLRMLGQDPLRQHHLLLOVCV 168

```

RESULT 12
 Q92257 PRELIMINARY; PRT; 113 AA.
 AC Q92257;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein RA0635.
 GN RA0635 OR SMAIL169.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Plasmid pSymba (megaplasmid 1).
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 Gurjal M., Hong A., Hultzer L., Hyman R.W., Kahn D., Kahn M.L.,
 Kallman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 "Nucleotide sequence and predicted functions of the entire
 Sinorhizobium meliloti pSymba megaplasmid."
 Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 RT EMBL; AE007253; AAK65293.1;
 DR Hypothetical protein; Plasmid; Complete proteome.
 KW SEQUENCE 113 AA; 12176 MW; A24DA206DFB92AC2 CRC64;

Query Match 7.1%; Score 63; DB 16; Length 113;
 Best Local Similarity 27.4%; Pred. No. 1e+02;
 Matches 23; Conservative 11; Mismatches 22; Indels 28; Gaps 4;

OY 64 DELLPEPMEALFSSSEGLKAR-----IFS-----FDGKDLRHPTPQKSVHGS 111
 DB 45 DEVLVTRKALDIAEPRLQOYIAVFSDIASIKFGDKELA----- 90
 OY 112 DPNGR-RLTESYCETWTEAPSAT 134
 DB 91 -PGGRVAVTADVSVMKVNHEPAT 113

RESULT 13
 Q31004 PRELIMINARY; PRT; 78 AA.
 ID Q31004;
 AC Q31004;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE MHC-g (Fragment).
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Cynthia, Lena, and Kasey;
 RX MEDLINE=96186524; PubMed=8606033;
 RA Castro M.J., Morales P., Fernandez-Soria V., Suarez B., Recio M.J.,
 Alvarez M., Martin-Villa M., Arnal-Villena A.;
 "Allelic diversity at the primate MHC-g locus: exon 3 bears stop
 codons in all cercopithecin sequences."
 Immunogenetics 43:327-336(1996).
 RT EMBL; U33291; AAB05936.1;
 DR InterPro; IPR001039; MHC_1.
 DR Pfam; PF00129; MHC_1; 1.
 DR PRINTS; PR01638; MHCCLASSI.
 DR PRODOM; PD000050; MHC_1; 1.
 FT NON_TER 1 78
 FT 1 78
 SO SEQUENCE 78 AA; 8766 MW; FF33FOA7B48EC7B3 CRC64;

Query Match 7.0%; Score 62.5; DB 7; Length 78;
 Best Local Similarity 26.8%; Pred. No. 71;
 Matches 22; Conservative 11; Mismatches 36; Indels 13; Gaps 3;

OY 77 GSEGPLKPGARIFSEDKVLRHPPTWPKSVHGSPPNRRRLTESYCETWTEAPSATGQ 136
 DB 5 GSDGRLLRLEYEQYAVDGNKYL--ALNEDLRSTFAD-----TMAQISRKCEANNAHQ 56
 OY 137 ASSLLGRLGOSASCHHAYT 158
 DB 57 RRAYLEG-----TVECELRHLY 73

RESULT 14
 Q9VAR4 PRELIMINARY; PRT; 116 AA.
 ID Q9VAR4;
 AC Q9VAR4;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE CG14519 protein.
 GN CG14519.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Laoko P.E., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of Drosophila melanogaster."
 Science 287:2185-2195(2000).
 RT EMBL; AE003767; AAF56837.1;
 DR FLYBase; FBgn0039619; CG14519.
 SO SEQUENCE 116 AA; 13377 MW; 01FB556FB41502D8 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:09:07 ; Search time 85 Seconds
(without alignments)
317.453 Million cell updates/sec

Title: US-09-171-607A-1

Perfect score: 893
Sequence: 1 VALNSPLSGMGRGIRGADPQ.....ASCHHAYIVLCIENSFMPTAS 170

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 831040

Minimum DB seq length: 0
Maximum DB seq length: 170

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:.*
1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	511	57.2	108	22	AAB49807	Murine endostatin
2	510	57.1	102	24	ABP97428	Human endostatin a
3	346	38.7	63	23	ABB08121	Human endostatin a
4	270	30.2	85	20	AAV22226	Apomigren. Homo s
5	270	30.2	85	22	AAV72941	Human anti-angioge
6	258	28.9	54	23	ABG31792	Human endostatin a
7	246	27.5	46	23	ABG31790	Human endostatin a
8	242.5	27.2	123	20	AAV29655	Human endostatin a
9	232	26.0	41	23	ABG31789	Human endostatin a

10	221	24.7	50	22	AAB35585	Antiangiogenic pen
11	217	24.3	45	22	AAB35586	Antiangiogenic pen
12	214.5	24.0	124	20	AAV29654	Mouse endostatin a
13	213	23.9	43	23	ABG31788	Human endostatin a
14	211	23.6	48	22	AAB49808	Murine endostatin
15	201	22.5	50	22	AAB35587	Antiangiogenic pen
16	189	21.2	44	23	ABG31787	Human endostatin a
17	189	21.2	44	23	ABG31791	Human endostatin a
18	182	20.4	31	23	ABB08122	Human endostatin a
19	160	17.9	35	20	AAV42232	Human endostatin a
20	160	17.9	35	20	AAV42247	Angiogenic homolog
21	158	17.7	31	22	AAB49876	Human endostatin p
22	152	17.0	35	22	AAB92435	Miscellaneous pept
23	147	16.5	30	18	AAW4651	Protein HF-COL-18
24	136	15.2	39	22	AAB35584	Antiangiogenic non
25	135	15.1	27	22	AAB49864	Human endostatin p
26	134	15.0	26	22	AAU00902	Human immunogenic
27	134	15.0	26	22	AAU97137	Human endostatin p
28	125	14.0	24	22	AAB49804	Human endostatin p
29	110	12.3	21	22	AAB49805	Human endostatin p
30	107	12.0	20	22	AAB49856	Human endostatin p
31	104	11.6	22	22	AAB49869	Human endostatin p
32	101	11.3	22	20	AAV42237	Human endostatin p
33	101	11.3	22	20	AAV43405	Multivalent ligand
34	96	10.8	25	22	AAB49866	Human endostatin p
35	94	10.5	16	22	ABB08064	Angiostatin-deri
36	94	10.5	16	22	AAB74258	Exemplary anti-ang
37	94	10.5	16	22	ABG97541	Human endostatin p
38	82	9.2	17	22	AAB49863	Human endostatin p
39	81	9.1	15	22	AAB49802	Human endostatin p
40	78	8.7	14	22	AAB49874	Human endostatin p
41	73	8.2	119	23	ABP08546	Human ORFX protein
42	70	7.8	155	20	AAV60570	Human normal bladd
43	67.5	7.6	106	22	AAO07953	Human polypeptide
44	67	7.5	11	22	AAB49861	Human endostatin p
45	67	7.5	11	22	AAB49872	Human endostatin p

ALIGNMENTS

RESULT 1
AAB49807
ID AAB49807 standard; Protein; 108 AA.
XX
AC AAB49807;
XX
DT 02-MAR-2001 (first entry)
XX
DE Murine endostatin peptide fragment SEQ ID NO: 20.
XX
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
KW cancer; inflammation; angiogenesis-dependent disease.
XX
XX
OS Mus musculus.
PN WO200067771-A1.
XX
PD 16-NOV-2000.
XX
PE 02-MAY-2000; 2000MO-US12063.
XX
PR 06-MAY-1999; 99US-0132907.
PR 14-JUL-1999; 99US-0353333.
PA (BURN-) BURNHAM INST.
XX
PI Vuori K;
XX
XX WPI; 2001-040937/05.
XX
XX Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and

PT diabetic retinopathy -
 XX
 PS Claim 10; Page 124; 146pp; English.
 CC The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of
 CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inopportune invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubeosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophilic joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.
 XX
 SO Sequence 108 AA;
 Query Match 57.2%; Score 511; DB 22; Length 108;
 Best Local Similarity 87.9%; Pred. No. 1.7e-54;
 Matches 94; Conservative 9; Mismatches 4; Indels 0; Gaps 0;
 OY 12 RGTCGAPCCQCARAVGLAGTFRALSSRLQDLYSYRRADRAVPIVNLKDELPPSW 71
 DB 1 RGTCGAPCCQCARAVGLAGTFRALSSRLQDLYSYRRADRAVPIVNLKDELPPSW 60
 OY 72 EALFSGSGPLPKPARIFSFQDKDYLRRHPWPQKSVHSGSPNGRRL 118
 DB 61 DSLFSGSGQGLQPGARIFSFQDKDYLRRHPWPQKSVHSGSPNGRRL 107
 RESULT 2
 ABP97428
 ID ABP97428 standard; Protein: 102 AA.
 XX
 AC ABP97428;
 XX
 DT 30-MAY-2003 (first entry)
 XX
 DE Human endostatin active fragment.
 XX
 KW Human; endostatin; vascular endothelial growth inhibitor;
 KW recombinant adenovirus; tumour; cancer; cytostatic; gene delivery;
 KW gene therapy.
 XX
 XX Homo sapiens.
 PH Key Location/Qualifiers
 FT MISC-difference 35 /note= "Encoded by GGC"
 FT MISC-difference 62 /note= "Encoded by GAG"
 FT MISC-difference 63 /note= "Encoded by AGC"
 XX
 XX CNI366040-A.
 XX
 XX 28-AUG-2002.
 XX
 XX 27-SEP-2001; 2001CN-0127894.
 XX
 XX 27-SEP-2001; 2001CN-0127894.
 XX
 XX (HUAN/) HUANG W.
 XX
 XX Huang W;
 XX
 XX WPI; 2003-240394/24.
 DR N-PSDB; AB275169.
 XX
 XX Recombinant virus of human vascular endothelial growth inhibitor -
 PT
 XX

PS Claim 1; Fig 1; 14pp; Chinese.
 XX
 CC The invention relates to a recombinant adenovirus comprising a human
 CC endostatin gene for direct expression in a eukaryotic cell. Endostatin
 CC is an inhibitor of vascular endothelial growth. The recombinant
 CC adenovirus can be used to deliver the endostatin gene to tumour cells
 CC where it can be expressed and inhibit tumour vascularisation. The present
 CC sequence represents an active portion of human endostatin which is
 CC encoded by an adenoviral vector of the invention.
 XX
 SO Sequence 102 AA;
 Query Match 57.1%; Score 510; DB 24; Length 102;
 Best Local Similarity 95.1%; Pred. No. 2e-54;
 Matches 97; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 OY 59 IVNLKDELPPSWALFSGSGPLPKPARIFSFQDKDYLRRHPWPQKSVHSGSPNGRRL 118
 DB 1 IVNLKDELPPSWALFSGSGPLPKPARIFSFQDKDYLRRHPWPQKSVHSGSPNGRRL 60
 OY 119 TESYCETWRTAPSPATGQASSLGGRLGQSAASHAYIVL 160
 DB 61 TWRYCETWRTAPSPATGQASSLGGRLGQSAASHAYIVL 102
 RESULT 3
 ABB08121
 ID ABB08121 standard; protein: 63 AA.
 XX
 AC ABB08121;
 XX
 DT 10-SEP-2002 (first entry)
 XX
 DE Human endostatin polypeptide fragment JCK-362 (residues 52-114).
 XX
 KW Endostatin; endothelial; cell proliferation; angiogenesis; cytostatic;
 KW antiarthritic; antiarteriosclerotic; ophthalmological; antidiabetic;
 KW anticancer; vulnery; gynaecological; human.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 33..63 /note= "JCK-367 peptide (ABB08122), specifically
 FT Modified-site 63 /note= "C-terminal amide"
 FT
 XX
 XX US2002058620-A1.
 XX
 XX 16-MAY-2002.
 XX
 XX 30-MAR-2001; 2001US-0822540.
 XX
 XX 03-APR-2000; 2000US-194561P.
 XX
 XX (PHOE-) PHOENIX PHARMACOLOGIES INC.
 XX
 XX Chang J;
 XX
 XX WPI; 2002-499519/53.
 XX
 XX New purified polypeptide useful for reducing tumor volume in an animal
 PT comprises a specified amino acid sequence
 XX
 XX Claim 1; Fig 1; 19pp; English.
 XX
 CC The invention relates to endostatin polypeptide fragments that can be
 CC used as inhibitors of endothelial cell proliferation and angiogenesis.
 CC The polypeptide inhibitors are useful for reducing tumour volume in an
 CC animal; for inhibiting the proliferation of endothelial cells in an
 CC animal; for targeting and binding purified polypeptide to an angiogenesis
 CC receptor; inhibiting metastasis, for identifying, isolating or purifying

be used to inhibit undesirable angiogenesis in a mammal. They can be used
for inhibition of endothelial activity such as endothelial cell
migration, inhibition of tumour growth, arrest of endothelial cells in G1
phase of the cell cycle, and inducing apoptosis in endothelial cells.
They can be used for treating e.g. angiogenesis-dependent cancers and
tumours, tumour metastasis, benign tumours e.g. haemangiomas, acoustic
neuromas, neurofibromas, trachomas, and pyogenic granulomas, rheumatoid
arthritis, psoriasis, ocular angiolemic diseases e.g. diabetic
retinopathy, retinopathy of prematurity, macular degeneration, corneal
graft rejection, neovascular glaucoma, retrorenal fibroplasia, rubeosis,
Osler-Webber syndrome, myocardial angiogenesis, plaque
neovascularisation, telangectasia, haemophilic joints, angiodysplasia,
wound granulation, intestinal adhesions, Crohn's disease,
atherosclerosis, scleroderma, hypertrophic scars i.e. keloids, or cat
scratch disease and ulcers, as a birth control agent by preventing
vasculatisation required for embryo implantation. They can also be used
for the production of antibodies. Using the methods, the AAs can be
produced in high yields.

Sequence 85 AA;

Query Match 30.2%; Score 270; DB 20; Length 85;
Best Local Similarity 58.0%; Pred No. 5e-25;
Matches 47; Conservative 13; Mismatches 21; Indels 0; Gaps 0

OY 88 IFSDGKDVLRHPTWPORSVWHGSDDPNRRLTESYCETWRTAPASATGOASSILGGRLIG 147
 |::|||::| : |||| ::||| :|:|||::| :| :||| :| :||
DB 1 IYSFGRIIDMTDPSMPQKYIMHGSSPHGVRLVDNVCENARTADTAVTGSLASTICKIID 60

OY 148 OSASCHHAIVLYLCIENSFPM 168
 | :| :| :| :|||||:
DB 61 OKAYSCANRLIVLCIENSFPMT 81

RESULT 5
AAAY72941
ID AAY72941 standard; peptide: 85 AA.
XX
AC AAY72941;
XX
DT 13-JUN-2001 (first entry)
XX
DE Human anti-angiogenic effector molecule, apomigren.
XX
KW Human; attenuated tumour-targeted bacteria; effector molecule;
KV Tumour necrosis factor; TNF; anti-angiogenic factor;
RW cytotoxic polypeptide; tumour inhibitory enzyme; therapy; carcinoma;
KM melanoma; lymphoma; sarcoma; metastasis; cytosolic; apomigren.
XX
OS Homo sapiens.
XX
PN WO200125397-A2.
XD 12-Apr-2001.
XX
PF 24-Aug-2000; 2000WO-US23242.
PR 04-Oct-1999; 99US-0157500.
PR 04-Oct-1999; 99US-0157581.
PR 04-Oct-1999; 99US-0157637.
XX
PA (VION-) VION PHARM INC.
PI Bermudes DG, King IC, Clairmont CA, Lin SL, Belcourt M;
DR WPI; 2001-245063/25.
PT Attenuated tumor-targeted bacteria comprising nucleic acids encoding
XS primary and secondary effector molecules, useful for the treatment of
XZ e.g. carcinomas, melanomas, lymphomas and sarcomas -
Example 14; Page 86; 185pp; English.

Query Match	Best Local Similarity	Score 270;	DB 22;	Length 85;
Matches	47; Conservative	13; Mismatches	21; Indels	0; Gaps
0Y	88 IFSPDGKDLRHPHPQKSVHMGSDPNGRRLTESYCEWTRTAPASATGQASSILGRLG	147		
Db	1 IYSEFGRIIMDPSPQKVIYMHGSSPHGRLVDNYCEAMRTADTAVTGLASPLSGKIID	60		
0Y	148 OSASCHHAYTYLCLIESEFMT	168		
Db	61 QKAYSCANRLIVLCLIESEFMT	81		
RESULT 6				
ABG31792				
ID	ABG31792 standard; Protein; 54 AA.			
XX	ABG31792;			
AC				
XX	05-DEC-2002 (first entry)			
DT				
XX	Human endostatin antlangiogenic peptide #6.			
DE				
XX	Human; endostatin; antlangiogenic peptide; tumour; cancer; metastasis;			
KW	cytostatic; antlangiogenic.			
RW				
XX	Homo sapiens.			
OS				
XX				
Key	Location/Qualifiers			
FN	1 /label= OTHER			
FN	/note= "OTHER- optionally tert-butoxycarbonyl-histidine"			
FT	7 /label= OTHER			
FT	/note= "OTHER- optionally Ser(tbu)"			
FT	10 /label= OTHER			
FT	/note= "OTHER- optionally Ser(tbu)"			
FT	14 /label= OTHER			
FT	/note= "OTHER- optionally Arg(Pbf)"			
FT	17 /label= OTHER			
FT	/note= "OTHER- optionally Arg(Pbf)"			
FT	20 /label= OTHER			
FT	/note= "OTHER- optionally Arg(Pbf)"			
FT	23 /label= OTHER			
FT	/note= "OTHER- optionally Asp(otbu)"			
FT	23 /label= OTHER			
FT	/note= "OTHER- optionally Cys(Acm)"			
FT	28 /label= OTHER			
FT	/note= "OTHER- optionally Arg(Pbf)"			
FT	35 /label= OTHER			
FT	/note= "OTHER- optionally Thr(tbu)"			

Accession	Protein	Score	Length	Matches	Conservative	Mismatches	Indels	Gaps
FT	Modified-site	37						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Arg(Pbf)"							
FT	Modified-site	41						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Ser(Lbu)"							
FT	Modified-site	42						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Ser(Lbu)"							
FT	Modified-site	43						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Arg(Pbf)"							
FT	Modified-site	46						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Asp(OCbu)"							
FT	Modified-site	48						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Tyr(Lbu)"							
FT	Modified-site	49						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Ser(Lbu)"							
FT	Modified-site	52						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Arg(Pbf)"							
FT	Modified-site	53						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Arg(Pbf)"							
FT	Modified-site	54						
FT	/label= OTHER							
FT	/note= "OTHER- Optionally Ala-resln"							
PN	WO200268457-A2.							
XX								
PD	06-SEP-2002.							
XX								
PF	27-FEB-2002; 2002MO-IT00119.							
XX								
PK	27-FEB-2001; 2001IT-MI00394.							
XX								
PA	(UYMI-) UNIV MILANO.							
PI	Chillemi F, Vicentini LMT, Francescato P;							
DR	WPI: 2002-698655/75.							
XX								
PT	New peptide useful for the preparation of medicaments with							
PT	antitumorigenic activity that may be used in treating tumours or							
PT	metastases; comprises a sequence corresponding to fragments of human							
PT	endostatin							
XX								
PS	Example 12; Page 12-13; 24pp; English.							
XX								
CC	The invention relates to peptide comprising 20-50 amino acids with							
CC	sequences corresponding to the human endostatin polypeptide sequence, its							
CC	salt or non-toxic derivative. The peptides are useful in the preparation							
CC	of medicaments with antitumorigenic activity which may be useful in							
CC	treating tumours or metastases. This sequence represents a human							
CC	endostatin antitumorigenic peptide of the invention.							
XX								
Sequence	54 AA:							
Query Match	28.9%; Score 258; DB 23; Length 54;							
Best Local Similarity	100.0%; Pred. No. 7.7e-24;							
Matches	52; Conservative 0; Mismatches 0; Indels 0; Gaps 0.							
QY	1 VALNSPLSGMGRIGADFCFOQARAVGLAGTFRFAFLSSRLQDLYSVRR 52							
DB	3 VALNSPLSGMGRIGADFCFOQARAVGLAGTFRFAFLSSRLQDLYSVRR 54							

```

XX ABG31790;
AC 05-DEC-2002 (first entry)
DT Human endostatin antilanglogenic peptide #4.
XX Human; endostatin; antilanglogenic peptide; tumour; cancer; metastasis;
KW cytosstatic; antilanglogenic.
XX Homo sapiens.
OS
FH Key
FH Modified-site
FT 1 Location/Qualifiers
FT /label= OTHER
FT /note= "OTHER- Optionally Tyr(tBu)"
FT Modified-site
FT 2 /label= OTHER
FT /note= "OTHER- Optionally tert-butoxycarbonyl-cysteine"
FT Modified-site
FT 3 /label= OTHER
FT /note= "OTHER- Optionally Glu(OtBu)"
FT Modified-site
FT 4 /label= OTHER
FT /note= "OTHER- Optionally Thr(tBu)"
FT Modified-site
FT 5 /label= OTHER
FT /note= "OTHER- Optionally tert-butoxycarbonyl-tryptophan"
FT Modified-site
FT 6 /label= OTHER
FT /note= "OTHER- Optionally Arg(Pbf)"
FT Modified-site
FT 7 /label= OTHER
FT /note= "OTHER- Optionally Thr(tBu)"
FT Modified-site
FT 8 /label= OTHER
FT /note= "OTHER- Optionally Glu(OtBu)"
FT Modified-site
FT 11 /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site
FT 13 /label= OTHER
FT /note= "OTHER- Optionally Thr(tBu)"
FT Modified-site
FT 17 /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site
FT 18 /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site
FT 23 /label= OTHER
FT /note= "OTHER- Optionally Arg(Pbf)"
FT Modified-site
FT 28 /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site
FT 31 /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site
FT 32 /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site
FT 33 /label= OTHER
FT /note= "OTHER- Optionally tert-butoxycarbonyl-cysteine"
FT Modified-site
FT 34 /label= OTHER
FT /note= "OTHER- Optionally tert-butoxycarbonyl-histidine"
FT Modified-site
FT 36 /label= OTHER
FT /note= "OTHER- Optionally tert-butoxycarbonyl-histidine"
FT Modified-site
FT 40 /label= OTHER
FT /note= "OTHER- Optionally Tyr(tBu)"
FT Modified-site
FT /label= OTHER
FT /note= "OTHER- Cys(tBu)"

```

```

FT Modified-site 42
FT /label= OTHER
FT /note= "OTHER- Optionally Glu(OtBu)"
FT Modified-site 44
FT /label= OTHER
FT /note= "OTHER- Optionally Ser(tBu)"
FT Modified-site 46
FT /label= OTHER
FT /note= "OTHER- Optionally Met-resin"
XX WO200268457-A2.
XX 06-SEP-2002.
XX 27-FEB-2002; 2002WO-IT00119.
XX 27-FEB-2001; 2001IT-MI00394.
XX (UWMI-) UNIV MILANO.
XX Chllem F, Vincetnle LMT, Francescato P;
XX WPI: 2002-698655/75.
XX
XX New peptide useful for the preparation of medicaments with
XX antilanglogenic activity that may be used in treating tumours or
XX metastases, comprises a sequence corresponding to fragments of human
XX endostatin
XX
XX Example 7; Page 10; 24pp; English.
XX
XX The invention relates to peptide comprising 20-50 amino acids with
XX sequences corresponding to the human endostatin polypeptide sequence, its
XX salt or non-toxic derivative. The peptides are useful in the preparation
XX of medicaments with antilanglogenic activity which may be useful in
XX treating tumours or metastases. This sequence represents a human
XX endostatin antilanglogenic peptide of the invention.
XX
XX Sequence 46 AA:
XX
XX Query Match 27.5%; Score 246; DB 23; Length 46;
XX Best Local Similarity 100.0%; Pred. No. 1 8e-22;
XX Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 122 YCETWRTDAPSATGQASSLLGGRLGQSAASCHHAYIVLCIENSFM 167
XX 1 YCETWRTDAPSATGQASSLLGGRLGQSAASCHHAYIVLCIENSFM 46
XX
XX RESULT 8
XX ID AAY29655 standard; Protein; 123 AA.
XX AC AAY29655;
XX DT 03-NOV-1999 (first entry)
XX DE Human endostatin amino acid sequence.
XX DE
XX KW Endostatin; inclusion body; expression; anticancer; antimetastatic;
XX anglogenesis; migration; proliferation; inhibition; collagen XVIII;
XX tumour.
XX OS Homo sapiens.
XX OS
XX PN WO9942486-A1.
XX PD 26-AUG-1999.
XX XX 19-FEB-1999; 99WO-US03271.
XX PR 23-FEB-1998; 98US-0075587.
XX

```

PA (SEAR) SEARLE & CO G D.
 XX
 PI Harding ET, Violand BN;
 XX
 DR WPI: 1999-527458/44.
 XX
 PT Production of endostatin by refolding recombinant protein, useful as
 PT anticancer and antimetastatic agent
 XX
 PS Disclosure; Page 28; 52pp; English.
 XX
 CC A method has been developed for the production of endostatin (I). The
 CC method comprises: (i) culturing host cells expressing the gene for (I);
 CC (ii) recovering the gene product; (iii) refolding it at near neutral pH;
 CC and (iv) isolating properly folded (I). (I) are inhibitors of
 CC angiogenesis, by inhibiting migration and proliferation of endothelial
 CC cells. The method is used to produce human or murine (I) from inclusion
 CC bodies formed in recombinant bacteria. (I), the C-terminal fragment of
 CC type alpha 1 collagen XVIII, is known to inhibit growth of some
 CC metastases and primary tumours. The present sequence represents the
 CC human endostatin amino acid sequence given in the present invention.
 XX
 SQ Sequence 123 AA;
 Query Match 27.2%; Score 242.5; DB 20; Length 123;
 Best Local Similarity 60.7%; Pred. No. 2.1e-21;
 Matches 99; Conservative 2; Mismatches 9; Indels 53; Gaps 26;
 OY 8 SGGKRGIGADPEQCEQARAVGLAGTFPAFLSRLODLYIVRRADRAAPVIVNKDELL 67
 DB 13 SGGHRC-RGAD-C---ARAVG-AGT-RA--SSR---DYS-VRRADRAAV--VN-KD-- 52
 OY 68 FPSWEALFSGSEGLKPGARIFSEFGKDYLRHPTMPQKSVHSGSDPNGRRLTESYCEIWR 127
 DB 53 --SWA---SGS---GKGAR---SDGKQV-RH-TW--KSVHGSQD-NGR--TSIC-TWR 92
 OY 128 TEASATGQASSLLGRLGQSAASHHAYIVLCIENSEPTAS 170
 DB 93 T--ASATGASS---GGR---GSAASHHAYV--C--NS-MTAS 122
 RESULT 9
 ABG31789
 ID ABG31789 standard; Protein; 41 AA.
 XX
 AC ABG31789;
 05-DEC-2002 (first entry)
 DE Human endostatin antiangiogenic peptide #3.
 XX
 KW Human; endostatin; antiangiogenic peptide; tumour; cancer; metastasis;
 KW cytosol; antiangiogenic.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Modified-site
 FT 3 Location/Qualifiers
 FT /label= OTHER
 FT /note= OTHER- Optionally tert-butoxycarbonyl-lysine"
 FT Modified-site
 FT 7 /label= OTHER
 FT /note= OTHER- Optionally Arg(Pbf)"
 FT Modified-site
 FT 10 /label= OTHER
 FT /note= OTHER- Optionally Ser(tbu)"
 FT Modified-site
 FT 12 /label= OTHER
 FT /note= OTHER- Optionally Asp(OtBu)"
 FT Modified-site
 FT 14 /label= OTHER
 FT /note= OTHER- Optionally tert-butoxycarbonyl-lysine"
 FT Modified-site
 FT 15 /label= OTHER- Optionally tert-butoxycarbonyl-lysine"

FT
 FT /label= OTHER
 FT /note= OTHER- Optionally Asp(OtBu)"
 FT Modified-site
 FT 18 /label= OTHER
 FT /note= OTHER- Optionally Arg(Pbf)"
 FT Modified-site
 FT 19 /label= OTHER
 FT /note= OTHER- Optionally trityl-histidine"
 FT Modified-site
 FT 21 /label= OTHER
 FT /note= OTHER- Optionally Thr(tbu)"
 FT Modified-site
 FT 22 /label= OTHER
 FT /note= OTHER- Optionally tert-butoxycarbonyl-tryptophan"
 FT Modified-site
 FT 25 /label= OTHER
 FT /note= OTHER- Optionally tert-butoxycarbonyl-lysine"
 FT Modified-site
 FT 26 /label= OTHER
 FT /note= OTHER- Optionally Ser(tbu)"
 FT Modified-site
 FT 28 /label= OTHER
 FT /note= OTHER- Optionally tert-butoxycarbonyl-tryptophan"
 FT Modified-site
 FT 29 /label= OTHER
 FT /note= OTHER- Optionally trityl-histidine"
 FT Modified-site
 FT 31 /label= OTHER
 FT /note= OTHER- Optionally Ser(tbu)"
 FT Modified-site
 FT 32 /label= OTHER
 FT /note= OTHER- Optionally Asp(OtBu)"
 FT Modified-site
 FT 36 /label= OTHER
 FT /note= OTHER- Optionally Arg(Pbf)"
 FT Modified-site
 FT 37 /label= OTHER
 FT /note= OTHER- Optionally Arg(Pbf)"
 FT Modified-site
 FT 39 /label= OTHER
 FT /note= OTHER- Optionally Thr(tbu)"
 FT Modified-site
 FT 40 /label= OTHER
 FT /note= OTHER- Optionally Glu(OtBu)"
 FT Modified-site
 FT 41 /label= OTHER
 FT /note= OTHER- Optionally Ser(tbu)-resin"
 FT
 XX
 PN WO200268457-A2.
 XX
 PD 06-SEP-2002.
 XX
 PP 27-FEB-2002; 2002WO-IT00119.
 XX
 PR 27-FEB-2001; 2001IT-MI00394.
 XX
 PA (UYMI-) UNIV MILANO.
 XX
 PI Chillemi F, Vicentini LMF, Francescato P;
 XX
 DR WPI: 2002-698655/75.
 XX
 PT New peptide useful for the preparation of medicaments with
 PT antiangiogenic activity that may be used in treating tumours or
 PT metastases, comprises a sequence corresponding to fragments of human
 PT endostatin
 XX
 PS Example 5; Page 9; 24pp; English.
 XX
 CC The invention relates to peptide comprising 20-50 amino acids with
 CC sequences corresponding to the human endostatin polypeptide sequence, its
 CC salt or non-toxic derivative. The peptides are useful in the preparation
 CC of medicaments with antiangiogenic activity which may be useful in

CC treating tumours or metastases. This sequence represents a human
CC endostatin antiangiogenic peptide of the invention.
XX
SQ Sequence 41 AA;

Query Match 26.0%; Score 232; DB 23; Length 41;
Best Local Similarity 100.0%; Pred. No. 8,1e-21;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 81 PLKPGARIFSGDKVLRHPTWPKSVWHSDDPNCGRRLTES 121
DB 1 PLKPGARIFSGDKVLRHPTWPKSVWHSDDPNCGRRLTES 41

RESULT 10
AAB35585
ID AAB35585 standard; peptide: 50 AA.
XX
AC AAB35585;

DT 14-FEB-2001 (first entry)

XX Antiangiogenic pentacontapeptide II.

XX Antiangiogenic; angiogenesis; cancer; endostatin.

OS Synthetic.

XX WO200063249-A1.

PD 26-OCT-2000.

XX 11-APR-2000; 2000WO-EP03236.

XX 15-APR-1999; 99IT-MI00777.

XX (UYMI-) UNIV MILANO.
XX (UYFI-) UNIV FIRENZE.

XX Chillemi F, Francescato P, Ziche M;

DR WPI: 2001-007005/01.

PT Polypeptides derived from endostatin exhibiting antiangiogenic activity
PT useful for treatment of angiogenesis-dependent tumours

PS Claim 3; Page 18; 28pp; English.

XX The present invention describes a number of peptides derived from
CC endostatin which exhibit antiangiogenic activity. These may be used in
the treatment of cancer. The present sequence is one of the peptides of
the invention.

XX Sequence 50 AA;

Query Match 24.7%; Score 221; DB 22; Length 50;
Best Local Similarity 86.0%; Pred. No. 2,4e-19;
Matches 43; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 28 VGLAGTRAPLSSRLQDLYSIVRRADRAVPYIVLKDLEFPSEALFSG 77
DB 1 VGLSGTRAPLSSRLQDLYSIVRRADRAVPYIVLKDLEFPSEALFSG 50

RESULT 11

AAB35586
ID AAB35586 standard; peptide: 45 AA.

XX AAB35586;

DT 14-FEB-2001 (first entry)

XX Antiangiogenic pentatetracontapeptide III.

XX Antiangiogenic; angiogenesis; cancer; endostatin.
KW
XX
OS Synthetic.

XX WO200063249-A1.

PD 26-OCT-2000.

XX 11-APR-2000; 2000WO-EP03236.

XX 15-APR-1999; 99IT-MI00777.

XX (UYMI-) UNIV MILANO.
XX (UYFI-) UNIV FIRENZE.

XX Chillemi F, Francescato P, Ziche M;

DR WPI: 2001-007005/01.

PT Polypeptides derived from endostatin exhibiting antiangiogenic activity
PT useful for treatment of angiogenesis-dependent tumours

PS Claim 4; Page 18; 28pp; English.

XX The present invention describes a number of peptides derived from
CC endostatin which exhibit antiangiogenic activity. These may be used in
the treatment of cancer. The present sequence is one of the peptides of
the invention.

XX Sequence 45 AA;

Query Match 24.3%; Score 217; DB 22; Length 45;
Best Local Similarity 82.2%; Pred. No. 6,5e-19;
Matches 37; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 78 SEGPLKPGARIFSGDKVLRHPTWPKSVWHSDDPNCGRRLTESY 122
DB 1 SGQVQPGARIFSGDKVLRHPTWPKSVWHSDDPNCGRRLTESY 45

RESULT 12

AAV29654
ID AAV29654 standard; Protein: 124 AA.

XX AAV29654;

DT 03-NOV-1999 (first entry)

DE Mouse endostatin amino acid sequence.

XX Endostatin; inclusion body; expression; anticancer; antimetastatic;
KW angiogenesis; migration; proliferation; inhibition; collagen XVIII;
tumour.

XX Mus sp.

XX WO9942486-A1.

XX 26-AUG-1999.

XX 19-FEB-1999; 99WO-US03271.

XX 23-FEB-1998; 98US-0075587.

XX (SEAR) SEARLE & CO G D.

XX Harding EI, Vieland BN;

DR WPI: 1999-527458/44.

PT Production of endostatin by refolding recombinant protein, useful as
antitumor and antimetastatic agent

XX Disclosure: Page 28; 52pp; English.
 PS A method has been developed for the production of endostatin (I). The
 CC method comprises: (i) culturing host cells expressing the gene for (I);
 CC (ii) recovering the gene product; (iii) refolding it at near neutral pH;
 CC and (iv) isolating properly folded (I). (I) are inhibitors of
 CC angiogenesis, by inhibiting migration and proliferation of endothelial
 CC cells. The method is used to produce human or murine (I) from inclusion
 CC bodies formed in recombinant bacteria. (I), the C-terminal fragment of
 CC type alpha 1 collagen XVIII, is known to inhibit growth of some
 CC metastases and primary tumours. The present sequence represents the
 CC mouse endostatin amino acid sequence given in the present invention.
 XX

SQ Sequence 124 AA:

Query Match 24.0%; Score 214.5; DB 20; Length 124;
 Best Local Similarity 52.8%; Pred. No. 5,9e-18;
 Matches 86; Conservative 13; Mismatches 13; Indels 51; Gaps 24;

8 SGGMNGIRGADFOCFQARAVGLAGTFRAFLSSRLQDLSYVRADRAAVPIVNLKDELL 67
 12 SGGMNG-RGAD--C--ARAVG-SGT-RA--SSR---DYS-VRRADRSV--VN-KD--- 51
 68 FPSWEALFSGSGBGLKPGARIFSEDGKDVLRHPWPQKSVHMGSDPNGRRLTESYCEWNR 127
 52 VSSMSSSGSG-----GAR---SDGRDV-RH-AW--KSVHMGSD-SCRRM--SYC-TWR 93
 128 TEAPSATGQASSLLGGRLLGOSASCHHAYIVLCTENSPMTRAS 170
 94 T-TTGATGASS-----GKRAKASCHNSYV--C--NS-MWSS 123

RESULT 13
 ID ABG31788 standard; Protein; 43 AA.
 XX
 AC ABG31788;
 XX
 DT 05-DEC-2002 (first entry)
 XX
 DE Human endostatin antiangiogenic peptide #2.
 XX
 KW Human; endostatin; antiangiogenic peptide; tumour; cancer; metastasis;
 KW cytostatic; antiangiogenic.
 XX
 XX Homo sapiens.
 XX
 Key Location/Qualifiers
 FT Modified-site 2 /label= OTHER
 FT /note= "OTHER- Optionally Ser(tbu)"
 FT Modified-site 3 /label= OTHER
 FT /note= "OTHER- Optionally Ser(tbu)"
 FT Modified-site 4 /label= OTHER
 FT /note= "OTHER- Optionally Arg(pbf)"
 FT Modified-site 7 /label= OTHER
 FT /note= "OTHER- Optionally Asp(Obu)"
 FT Modified-site 9 /label= OTHER
 FT /note= "OTHER- Optionally Tyr(tbu)"
 FT Modified-site 10 /label= OTHER
 FT /note= "OTHER- Optionally Ser(tbu)"
 FT Modified-site 13 /label= OTHER
 FT /note= "OTHER- Optionally Arg(pbf)"
 FT Modified-site 14 /label= OTHER
 FT /note= "OTHER- Optionally Arg(pbf)"
 FT Modified-site 14 /label= OTHER
 FT /note= "OTHER- Optionally Arg(pbf)"

FT Modified-site 16 /label= OTHER
 FT /note= "OTHER- Optionally Asp(Obu)"
 FT Modified-site 17 /label= OTHER
 FT /note= "OTHER- Optionally Arg(pbf)"
 FT Modified-site 26 /label= OTHER
 FT /note= "OTHER- Optionally tert-butoxycarbonyl-lysine"
 FT Modified-site 27 /label= OTHER
 FT /note= "OTHER- Optionally Asp(Obu)"
 FT Modified-site 28 /label= OTHER
 FT /note= "OTHER- Optionally Gly(Obu)"
 FT Modified-site 33 /label= OTHER
 FT /note= "OTHER- Optionally Ser(tbu)"
 FT Modified-site 34 /label= OTHER
 FT /note= "OTHER- Optionally tert-butoxycarbonyl-tryptophan"
 FT Modified-site 35 /label= OTHER
 FT /note= "OTHER- Optionally Gly(Obu)"
 FT Modified-site 39 /label= OTHER
 FT /note= "OTHER- Optionally Ser(tbu)"
 FT Modified-site 41 /label= OTHER
 FT /note= "OTHER- Optionally Gly(Obu)"
 FT Modified-site 42 /label= OTHER
 FT /note= "OTHER- Optionally Ser(tbu)"
 FT Modified-site 43 /label= OTHER
 FT /note= "OTHER- Optionally Gly-resin"
 FT Modified-site 43 /label= OTHER
 FT /note= "OTHER- Optionally Gly-resin"

MO200268457-A2.
 XX
 PN 06-SEP-2002.
 XX
 PD 27-FEB-2002; 2002WO-IT00119.
 XX
 PR 27-FEB-2001; 2001IT-MI00394.
 XX
 PA (UYMI-) UNIV MILANO.
 XX
 PI Chillemi F, Vinciguerra LMT, Francescato P;
 XX WPI; 2002-698655/75.
 XX
 DR New peptide useful for the preparation of medicaments with
 XX antiangiogenic activity that may be used in treating tumours or
 PT metastases, comprises a sequence corresponding to fragments of human
 PT endostatin
 XX
 PS Example 3; Page 8; 24pp; English.
 XX
 CC The invention relates to peptide comprising 20-50 amino acids with
 CC sequences corresponding to the human endostatin polypeptide sequence, its
 CC salt or non-toxic derivative. The peptides are useful in the preparation
 CC of medicaments with antiangiogenic activity which may be useful in
 CC treating tumours or metastases. This sequence represents a human
 CC endostatin antiangiogenic peptide of the invention.
 XX
 SQ Sequence 43 AA:
 XX

Query Match 23.9%; Score 213; DB 23; Length 43;
 Best Local Similarity 100.0%; Pred. No. 1,9e-18;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

38 LSSRLQDLSYVRADRAAVPIVNLKDELLFPSWEALFSGSEG 80
 ||||||||||||||||||||||||||||||||||||||||



XX WPI: 2000-365617/31.
DR N-PSDB: AAA27005.

XX Novel endostatin capable of inhibiting endothelial cell proliferation
XX and angiogenesis, useful for treating angiogenesis-dependent cancers
PT and as birth control agents -
XX

PS Claim 3; Page 39; 68pp; English.

XX The present sequence is an alternate functional endostatin
CC protein. When the human endostatin gene sequence AAA27004 is
CC recombinantly expressed, an observable doublet of protein results, both
CC versions of which are functional endostatin proteins. The present
CC endostatin variant is the same as the protein encoded by AAA27004 minus
CC the first four amino acids. Recombinant mouse endostatin (20 mg/kg) was
CC administered subcutaneously to mice implanted with Lewis lung carcinomas.
CC There was tumour mass regression non-detectable levels after 12 days of
CC therapy due to the angiogenesis inhibitory activity of endostatin. Thus
CC the protein is useful for treatment of angiogenesis-dependent cancers.
CC The polynucleotide and polypeptide sequences of this endostatin are
CC useful for treating and diagnosis of tumours, ocular angiogenic
CC diseases, Osler-Weber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma
CC and wound granulation, for treatment of diseases related to excessive or
CC abnormal stimulation of endothelial cells e.g. intestinal adhesions,
CC atherosclerosis, scleroderma. The protein may also be useful as a birth
CC control agent by reducing or preventing uterine vascularisation. The
CC gene for endostatin may be isolated from cells or tissue that express
CC high levels of endostatin, eg. tumour cells, by generating cDNA from
CC mRNA using reverse transcriptase and then amplifying the DNA sequence.

XX Sequence 178 AA:

Query Match 100.0%; Score 893; DB 21; Length 178;

Best Local Similarity 100.0%; Pred. No. 3.7e-101; Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFAPFLSSRLQDLYSTVRRADRAAVTV 60
DB 9 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFAPFLSSRLQDLYSTVRRADRAAVTV 68
OY 61 NLKDELLPSPWEALFSGSEGLPKPGARIFSPDGKVDLHHPWPQKSVHSGDPNCRRLTE 120
DB 69 NLKDELLPSPWEALFSGSEGLPKPGARIFSPDGKVDLHHPWPQKSVHSGDPNCRRLTE 128
OY 121 SYCETWRTAPASATGOASSLLGRLGSAASCHHAYIVLCLENSEFMTAS 170
DB 129 SYCETWRTAPASATGOASSLLGRLGSAASCHHAYIVLCLENSEFMTAS 178

RESULT 2

AAU00900 ID AU000900 standard; Protein; 178 AA.

AC AAU00900;

XX 04-JUL-2001 (first entry)

DE Human Endostatin(TM) N-terminal deletion mutant protein#2.

XX Human; Endostatin(TM): angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophilic joint; angiofibroma; wound granulation; variant;
KW mutant; multin.

XX Homo sapiens.

OS AAU00901

XX WO200119989-A2.

XX 22-MAR-2001.

XX 14-SEP-2000; 2000MO-US25166.

XX 14-SEP-1999; 99US-0153698.

XX (ENTR-) ENTREMED INC.

XX Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
PI Bernejo LL, Mistry FR, Shepard SR, Schrimsher JL;

XX WPI: 2001-244802/25.

XX N-PSDB: AAS00868.

PT Producing Endostatin protein for treating angiogenesis mediated
PT diseases such as solid tumours, comprises recombinantly producing the
PT protein using an expression system, and recovering and purifying the
PT protein -

PS Claim 5; Page 33; 67pp; English.

XX The sequence represents Human Endostatin(TM) N-terminal deletion mutant
CC protein lacking the N-terminal 4 amino acids and the C-terminal lysine,
CC a natural variant recovered from fermentations of *Pichia pastoris*
CC cultures harbouring a expression plasmid containing the Endostatin(TM)
CC DNA sequence given in AAS00868. The new method of the invention is
CC useful for producing, recovering and purifying Endostatin (TM) from
CC biological sources, such as biological fluids, tissues, cells, culture
CC media, and fermentation media. Endostatin(TM) is useful for treating
CC angiogenesis mediated diseases such as solid tumours, blood borne
CC tumours, leukemias, tumour metastases, benign tumours, e.g. haemangioma,
CC acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas,
CC rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
CC e.g., diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, colon
CC cancer, retrolental fibroplasia, rubeosis, Osler-Weber Syndrome,
CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,
CC haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM)
CC is also useful for treating disease of excessive or abnormal stimulation
CC of endothelial cells such as intestinal adhesions, atherosclerosis,
CC scleroderma and hypertrophic scars. Higher yields of more purified, and
CC biologically active Endostatin(TM) are obtained by the new method.
CC Endostatin(TM) can be stored in buffers for extended periods of time, and
CC also subjected to lyophilisation, while preserving biological activity,
CC centrifugation of broth from fermentation steps in production is avoided,
CC preventing unwanted potential cellular lysis and contamination with
CC additional proteins, pigments, enzymes and other cellular chemicals and
CC debris.

XX Sequence 178 AA:

Query Match 100.0%; Score 893; DB 22; Length 178;

Best Local Similarity 100.0%; Pred. No. 3.7e-101; Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFAPFLSSRLQDLYSTVRRADRAAVTV 60
DB 9 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFAPFLSSRLQDLYSTVRRADRAAVTV 68
OY 61 NLKDELLPSPWEALFSGSEGLPKPGARIFSPDGKVDLHHPWPQKSVHSGDPNCRRLTE 120
DB 69 NLKDELLPSPWEALFSGSEGLPKPGARIFSPDGKVDLHHPWPQKSVHSGDPNCRRLTE 128
OY 121 SYCETWRTAPASATGOASSLLGRLGSAASCHHAYIVLCLENSEFMTAS 170
DB 129 SYCETWRTAPASATGOASSLLGRLGSAASCHHAYIVLCLENSEFMTAS 178

RESULT 3

AAU00901

ID AAU0901 standard; Protein: 179 AA.
 AC AAU0901;
 DT 04-JUL-2001 (first entry)
 DE Human Endostatin(TM) N-terminal mutant protein1.
 KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
 KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
 KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubecosis;
 KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
 KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 KW retinopathy of prematurity; macular corneal graft rejection;
 KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;
 KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 KW haemophilic joint; angiofibroma; wound granulation; mutant; mutein.
 OS Homo sapiens.
 XX MO200119989-A2.
 XX 22-MAR-2001.
 XX 14-SEP-2000; 2000WO-US25166.
 XX 14-SEP-1999; 99US-0153698.
 XX (ENTR-) ENTREMED INC.
 XX Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
 PI Bermejo LJ, Mistry FR, Shepard SR, Schrimsher JL;
 DR N-PSDB: AAS00868.
 DR WPI: 2001-244802/25.
 XX Producing Endostatin protein for treating angiogenesis mediated
 PT diseases such as solid tumours, comprises recombinantly producing the
 PT protein using an expression system, and recovering and purifying the
 PT protein.
 PS Claim 5; Page 32; 67pp; English.
 XX The sequence represents a Human Endostatin(TM) N-terminal deletion
 CC mutant lacking the N-terminal 4 amino acids. The new method of the
 CC invention is useful for producing, recovering and purifying Endostatin
 CC (TM) from biological sources, such as biological fluids, tissues, cells,
 CC culture media, and fermentation media. Endostatin(TM) is useful for
 CC treating angiogenesis mediated diseases such as solid tumours, blood
 CC borne tumours, leukaemias, tumour metastases, benign tumours, e.g.
 CC haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
 CC e.g., diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, colon
 CC cancer, retrolental fibroplasia, rubecosis, Osler-Weber Syndrome,
 CC myocardioid angiogenesis, plaque neovascularisation, telangiectasia,
 CC haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM)
 CC is also useful for treating disease of excessive or abnormal stimulation
 CC of endothelial cells such as intestinal adhesions, atherosclerosis,
 CC scleroderma and hypertrophic scars. Higher yields of more purified, and
 CC biologically active Endostatin(TM) are obtained by the new method.
 CC Endostatin(TM) can be stored in buffers for extended periods of time, and
 CC also subjected to lyophilisation, while preserving biological activity.
 CC Centrifugation of broth from fermentation steps in production is avoided,
 CC preventing unwanted potential cellular lysis and contamination with
 CC additional proteins, pigments, enzymes and other cellular chemicals and
 CC debris.
 XX Sequence 179 AA;
 S0

QY 1 VALNSPLSGMKGIRGADFCFOQARAVGLACTFPRAFLSSRLQDLXSYRRADRAAPVY 60
 DB 9 VALNSPLSGMKGIRGADFCFOQARAVGLACTFPRAFLSSRLQDLXSYRRADRAAPVY 68
 QY 61 NLKDELLFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSVHSGDPNGRRRLTE 120
 DB 69 NLKDELLFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSVHSGDPNGRRRLTE 128
 QY 121 SYCETWRTPSPATQASSLLGRLIGQSAASCHNAVYVLCTIENSFMTAS 170
 DB 129 SYCETWRTPSPATQASSLLGRLIGQSAASCHNAVYVLCTIENSFMTAS 178
 RESULT 4
 ID AAB28399 standard; Protein: 182 AA.
 AC AAB28399;
 XX 19-FEB-2001 (first entry)
 DE Human endostatin.
 KW Human; endostatin; cytosstatic; antiproliferative;
 KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
 KW cancer; vascularised solid tumour.
 OS Homo sapiens.
 XX MO200064946-A2.
 XX 02-NOV-2000.
 XX 28-APR-2000; 2000WO-US11367.
 XX 28-APR-1999; 99US-0131432.
 XX (TEXA) UNIV TEXAS SYSTEM.
 XX Thorpe PE, Brekken RA;
 DR WPI: 2000-687317/67.
 PT Immunogenic composition for the treatment and diagnosis of cancer
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
 PS Example 10; Page 291-292; 298pp; English.
 XX The present invention relates to anti-Vascular Endothelial Growth Factor
 CC (VEGF) antibodies that bind to the same epitope as the monoclonal
 CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
 CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
 CC receptor VEGFR1. The present sequence is human endostatin. Endostatin
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.
 CC The anti-VEGF antibodies of the present invention are useful for the
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.
 XX Sequence 182 AA;
 S0

Query Match 100.0%; Score 893; DB 21; Length 182;
 Best Local Similarity 100.0%; Pred. No. 3.8e-101;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMKGIRGADFCFOQARAVGLACTFPRAFLSSRLQDLXSYRRADRAAPVY 60
 DB 13 VALNSPLSGMKGIRGADFCFOQARAVGLACTFPRAFLSSRLQDLXSYRRADRAAPVY 72
 QY 61 NLKDELLFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSVHSGDPNGRRRLTE 120
 DB 73 NLKDELLFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKSVHSGDPNGRRRLTE 132

OY 121 SYCETWTEAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMTAS 170
DB 133 SYCETWTEAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMTAS 182

RESULT 5
AAV94323

ID AAV94323 standard; Protein: 182 AA.

AC AAV94323;

DT 11-AUG-2000 (first entry)

DE Human endostatin protein.

KW Human: endothelial cell proliferation inhibitor; collagen XVIII;
KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;
KW vasotrophic; dermatological; ophthalmological; vulvar; artery;
KW antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW ocular angiogenic disease; atherosclerosis; scleroderma;
KW myocardial angiogenesis; telangiectasia; angiofibroma;
KW wound granulation.

OS Homo sapiens.

PN MO200026368-A2.

PD 11-MAY-2000.

PE 01-NOV-1999; 99WO-US25605.

PR 30-OCT-1998; 98US-0106343.

PR 20-MAY-1999; 99US-0315689.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI O'Reilly MS, Folkman MJ;

DR WPI: 2000-365617/31.

DR N-PSDB; AAA27004.

PT Novel endostatin capable of inhibiting endothelial cell proliferation

PT and angiogenesis, useful for treating angiogenesis-dependent cancers

PT and as birth control agents

PS Claim 2; Page 38; 68pp; English.

XX The present sequence is an endostatin protein which is the carboxy
XX terminal protein of human collagen XVIII. Recombinant mouse endostatin
XX (20 mg/kg) was administered subcutaneously to mice implanted with Lewis
XX lung carcinomas. There was tumour mass regression non-detectable levels
XX after 12 days of therapy due to the angiogenesis inhibitory activity of
XX endostatin. Thus the protein is useful for treatment of angiogenesis-
XX dependent cancers. The polynucleotide and polypeptide sequences of this
XX endostatin are useful for treating and diagnosis of tumours, ocular
XX angiogenic diseases, Osler-Webber syndrome, myocardial angiogenesis,
XX plaque neovascularisation, telangiectasia, haemophilic joints,
XX angiofibroma and wound granulation, for treatment of diseases related to
XX adhesions, atherosclerosis, scleroderma. The protein may also be useful
XX as a birth control agent by reducing or preventing uterine
XX vascularisation. The gene for endostatin may be isolated from cells or
XX tissue that express high levels of endostatin, eg. tumour cells, by
XX generating cDNA from mRNA using reverse transcriptase and then amplifying
XX the DNA sequence.

SO Sequence 182 AA;

Query Match 100.0%; Score 893; DB 21; Length 182;

Best Local Similarity 100.0%; Pred. No. 3.8e-101;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YALNSPLSGMGKIRGADRFQCFQOARAVGLAGTFRAFLSLRLDLYSVRRADRAVPIV 60

DB 13 VALNSPLSGMGKIRGADRFQCFQOARAVGLAGTFRAFLSLRLDLYSVRRADRAVPIV 72

OY 61 NLKDELLFPSSWEALFSGSEGLPKGARIFSPDGKDVLRHPWPQKSVHSGDPNCRRLTE 120

DB 73 NLKDELLFPSSWEALFSGSEGLPKGARIFSPDGKDVLRHPWPQKSVHSGDPNCRRLTE 132

OY 121 SYCETWTEAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMTAS 170
DB 133 SYCETWTEAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMTAS 182

RESULT 6
AAV59622

ID AAV59622 standard; Protein: 182 AA.

AC AAV59622;

DT 14-MAR-2000 (first entry)

DE Human endostatin protein fragment.

KW Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;
KW metastatic cancer; tumorigenesis; ocular angiogenic disease;
KW rheumatoid arthritis; Osler-Webber syndrome; telangiectasia;
KW haemophilic joint; angiofibroma; wound granulation.

OS Homo sapiens.

PN WO9962944-A2.

PD 09-DEC-1999.

PE 03-JUN-1999; 99WO-US12278.

PR 03-JUN-1998; 98US-0087890.

PR 10-JUL-1998; 98US-0092393.

PR 01-SEP-1998; 98US-0098790.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI Javaherian K, Folkman MJ;

DR WPI: 2000-072833/06.

PT New endostatin oligomers, used for treating e.g. tumours -

PT Disclosure; Page 6; 44pp; English.

XX This sequence is a fragment of the human endostatin protein. Endostatin
XX is an approximately 20KD C-terminal globular domain of the collagen-like
XX protein collagen XVIII. Protein oligomers consisting of more than one
XX endostatin monomer have anti-tubulogenic effects and induce
XX reorganization of the actin cytoskeleton. The oligomer has scatter factor
XX activity. The oligomers induce the destruction of tubular lumens and
XX elongation of cells, and inhibit tubulogenesis and tumorigenesis. The
XX oligomers can also be used to treat metastatic cancers, tumours,
XX rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Webber
XX syndrome, plaque neovascularisation, telangiectasia, haemophilic
XX joints, angiofibroma and wound granulation. The oligomers can also be
XX used to treat diseases that have angiogenesis as a pathological
XX consequence e.g. ulcers. The endostatin oligomers can also be used to
XX develop affinity columns for isolating antibodies or receptors. Passive
XX antibody therapy using antibodies that specifically bind endostatin
XX oligomers can be used to modulate morphogenic processes such as
XX metastatic cancer as well as angiogenesis-dependent processes such as
XX reproduction, development, wound healing, tissue repair, and
XX angiogenesis-dependent diseases. Also, antisera directed to the Fab
XX regions of endostatin oligomer antibodies can be administered to block
XX the ability of endogenous endostatin oligomer antisera to bind endostatin
XX oligomers.

SO Sequence 182 AA;

Query Match 100.0%; Score 893; DB 21; Length 182;
 Best Local Similarity 100.0%; Pred. No. 3.8e-101;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAIVPV 60
 DB 13 VALNSPLSGMGRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAIVPV 72
 OY 61 NLKDELLFPSPWEALFSGSEGPLKPGARIFSPDGKDYLRHPWPPOKSVWHGSDPNGRRLTE 120
 DB 73 NLKDELLFPSPWEALFSGSEGPLKPGARIFSPDGKDYLRHPWPPOKSVWHGSDPNGRRLTE 132
 OY 121 SYCETWTEAPSATGQASSLLGGRLLGQSAASHHAYIVLCIENSFMTAS 170
 DB 133 SYCETWTEAPSATGQASSLLGGRLLGQSAASHHAYIVLCIENSFMTAS 182

RESULT 7
 AAU00897

AAU00897 standard; Protein: 182 AA.

AAU00897;

04-JUL-2001 (first entry)

Human Endostatin(TM) C-terminus minus 1 protein.

Human: Endostatin(TM): angiogenesis mediated disease; solid tumours;
 blood borne tumour; leukaemia; tumour metastasis; benign tumour;
 haemangioma; acoustic neuroma; neurofibroma; trachoma; rubecosis;
 pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
 ocular angiogenic disease; diabetic retinopathy; macular degeneration;
 retinopathy of prematurity; macular corneal graft rejection;
 neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
 myocardial angiogenesis; plaque neovascularisation; telangiectasia;
 hemophilic joint; angiodioma; wound granulation; variant;
 C-terminus minus 1 protein.

Homo sapiens.

WC200119989-A2.

22-MAR-2001.

14-SEP-2000: 2000MO-US25166.

14-SEP-1999: 99US-0153698.

(ENTR-) ENTREMED INC.

Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;
 Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;

WPI: 2001-244802/25.
 N-PSDB: AAS00897.

Producing Endostatin protein for treating angiogenesis mediated
 diseases such as solid tumours, comprises recombinantly producing the
 protein using an expression system, and recovering and purifying the
 protein

Claim 5; Page 30; 67pp; English.

The sequence represents Human Endostatin(TM) C-terminus minus 1
 protein, a natural variant lacking the C-terminal amino acid of
 Endostatin(TM) recovered from fermentations of *Pichia*
 pastoris cultures harbouring a expression plasmid containing the
 Endostatin(TM) DNA sequence given in AAS00867. The new method of the
 invention is useful for producing, recovering and purifying Endostatin
 (TM) from biological sources, such as biological fluids, tissues, cells,
 culture media, and fermentation media. Endostatin(TM) is useful for
 treating angiogenesis mediated diseases such as solid tumours, blood

CC borne tumours, leukaemias, tumour metastases, benign tumours, e.g.
 CC haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
 CC e.g., diabetic retinopathy, retinopathy of prematurity, macular
 CC degeneration, corneal graft rejection, neovascular glaucoma, colon
 CC cancer, retrolental fibroplasia, rubecosis, Osler-Webber Syndrome,
 CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,
 CC haemophilic joints, angiodioma, and wound granulation. Endostatin(TM)
 CC is also useful for treating disease of excessive or abnormal stimulation
 CC of endothelial cells such as intestinal adhesions, atherosclerosis,
 CC scleroderma and hypertrophic scars. Higher yields of more purified, and
 CC biologically active Endostatin(TM) are obtained by the new method.
 CC Endostatin(TM) can be stored in buffers for extended periods of time, and
 CC also subjected to lyophilisation, while preserving biological activity.
 CC Centrifugation of broth from fermentation steps in production is avoided,
 CC preventing unwanted potential cellular lysis and contamination with
 CC additional proteins, pigments, enzymes and other cellular chemicals and
 CC debris.

Sequence 182 AA:

Query Match 100.0%; Score 893; DB 22; Length 182;
 Best Local Similarity 100.0%; Pred. No. 3.8e-101;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VALNSPLSGMGRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAIVPV 60
 DB 13 VALNSPLSGMGRGIRGADFCFOQARAVGLAGTFRAFLSSRLQDLYSIVRRADRAAIVPV 72
 OY 61 NLKDELLFPSPWEALFSGSEGPLKPGARIFSPDGKDYLRHPWPPOKSVWHGSDPNGRRLTE 120
 DB 73 NLKDELLFPSPWEALFSGSEGPLKPGARIFSPDGKDYLRHPWPPOKSVWHGSDPNGRRLTE 132
 OY 121 SYCETWTEAPSATGQASSLLGGRLLGQSAASHHAYIVLCIENSFMTAS 170
 DB 133 SYCETWTEAPSATGQASSLLGGRLLGQSAASHHAYIVLCIENSFMTAS 182

RESULT 8
 AAU77951

AAU77951 standard; Protein: 182 AA.

AAU77951;

02-JUL-2002 (first entry)

Amino acid sequence for human endostatin.

Human: immunocjugate; anti-vascular endothelial growth factor antibody;
 anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;
 KW VEGFR2; KDR/Flk-1; VEGFR1; Flt-1; angiogenesis; macular degeneration;
 KW ocular neovascular disease; cancer; vascularised solid tumour; AIDS;
 KW metastatic tumour; endothelial cell proliferation; inflammatory disorder;
 KW atherosclerosis; diabetic retinopathy; corneal graft rejection;
 KW acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;
 KW sickle cell anaemia; endometriosis; endostatin.

Homo sapiens.

AU200179401-A.

06-DEC-2001.

12-OCT-2001: 2001AU-0079401.

28-APR-2000: 2000AU-0048049.

12-OCT-2001: 2001AU-0079401.

(TEXA) UNIV TEXAS SYSTEM.

Thorpe PE, Brekken RA;
 WPI: 2002-281368/33.

XX Immunocjugate compositions for treating cancer by inhibiting
 PT angiogenesis and for delivering a diagnostic agent to tumour. comprises
 PT anti-vascular endothelial growth factor antibody attached to a
 PT biological agent

PS Example 10: Page 12-13 (Sequence listing): 300pp; English.

CC The present invention relates to antibody-based compositions comprising
 CC immunocjugate such as anti-vascular endothelial growth factor
 CC (VEGF) antibody (Ab) (or its antigen-binding fragment), attached to a
 CC biological agent, where the Ab binds to the same epitope as the
 CC monoclonal antibody (MAb) 2C3 ATCC PTA 1595, and significantly inhibits
 CC VEGF binding to the VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting
 CC of the invention are useful in therapy, and diagnosis, for inhibiting
 CC angiogenesis in an animal having ocular neovascular disease or macular
 CC degeneration, and for delivering a biological agent to a vascularised
 CC tumour. The compositions can also be used for treating cancer and
 CC subjects at risk of developing, a vascularised solid tumour, a metastatic
 CC tumour or metastases from a primary tumour. The composition is useful
 CC for specifically inhibiting VEGF-induced endothelial cell proliferation,
 CC without significantly inhibiting VEGF-induced macrophage, osteoclast or
 CC chondroclast function. The compositions can be used for treating various
 CC diseases such as inflammatory disorders, atherosclerosis, diabetic
 CC retinopathy, restenosis, acquired immune deficiency syndrome (AIDS),
 CC blood borne tumours, corneal graft rejection, Crohn's disease, fungal
 CC ulcers, infections, sickle cell anaemia, and endometriosis. The present
 CC sequence represents human endostatin. Endostatin may be attached or
 CC functionally associated with anti-VEGF antibodies.

SO Sequence 182 AA:

Query Match 100.0%; Score 893; DB 23; Length 182;
 Best Local Similarity 100.0%; Pred. No. 3.8e-101;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFRAFLLSRLODLYSTVRADRAAIVY 60
 DB 13 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFRAFLLSRLODLYSTVRADRAAIVY 72
 QY 61 NLKDELLFPSSWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVHSGSDPNGRRLTE 120
 DB 73 NLKDELLFPSSWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVHSGSDPNGRRLTE 132
 QY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
 DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 9

AA08693
 ID AAY08693 standard; Protein: 183 AA.

AA08693:

10-AUG-1999 (first entry)

Human endostatin protein fragment.

KW Plasmalogen; human; angiotatin; endostatin; gene therapy; vector;
 KW anti-angiogenic; attenuation; cytoskeletal; anti-diabetic; ophthalmology;
 KW tumour growth; solid tumour; diabetic retinopathy; retina.

OS Homo sapiens.

PN M09926480-A1.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-US24950.

XX 20-NOV-1997; 97US-0975424.

XX (GENE-) GENETIX PHARM INC.
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Bachelot T, Leboulch P, Pawluk RJ;

XX WPI; 1999-357696/30.

XX N-PSDB; AAX77719.

PT Anti-angiogenic gene therapy vectors

PS Disclosure: Page 74-75; 83pp; English.

CC This invention describes a novel viral gene therapy vector comprising a
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
 CC from human or murine endostatin, human or murine endostatin and
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
 CC sufficiently attenuated for use in human gene therapy. The products of
 CC the invention have anti-angiogenic, cytoskeletal, anti-diabetic and
 CC ophthalmological activity. The vector is used in gene therapy for
 CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
 CC expresses an anti-angiogenic polypeptide. An additional use comprises
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
 CC inhibits angiogenesis in the vicinity of the retina. The vector is
 CC administered to cells ex vivo and then administered to the patient.

SO Sequence 183 AA:

Query Match 100.0%; Score 893; DB 20; Length 183;
 Best Local Similarity 100.0%; Pred. No. 3.9e-101;
 Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFRAFLLSRLODLYSTVRADRAAIVY 60
 DB 13 VALNSPLSGMKGIRGADFOCFQOARAVGLAGTFRAFLLSRLODLYSTVRADRAAIVY 72
 QY 61 NLKDELLFPSSWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVHSGSDPNGRRLTE 120
 DB 73 NLKDELLFPSSWEALFSSGSEGLPKPGARIFSPDGKDVLRHPMPQKSVHSGSDPNGRRLTE 132
 QY 121 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 170
 DB 133 SYCETWRTAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFMTAS 182

RESULT 10

AA02113
 ID AAY02113 standard; Protein: 183 AA.

AA02113:

16-JUL-1999 (first entry)

SEQ ID 76 of W09916889.

Angiotatin; endostatin; interferon; thrombospondin;

KW anti-tumor; multifunctional protein; platelet factor 4; anti-angiogenic;

KW cancer; diabetic retinopathy; macular degeneration; arthritis;

KW tumor cell production.

OS Homo sapiens.

PN M09916889-A1.

XX 08-APR-1999.

XX 30-SEP-1998; 98WO-US20464.

XX 01-OCT-1997; 97US-0060609.

XX (SEAR) SEARLE & CO G D.

XX WPI: 2000-412290/35.

DR New angiogenesis-inhibiting protein receptors, useful in methods for
XX treating diseases and processes that are mediated by angiogenesis, such
PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -
XX
XX
XX Disclosure: Figure 3; 100pp; English.

XX This invention relates to angiogenesis-inhibiting protein receptors, and
CC the DNA sequences encoding them. Angiogenesis is the generation of new
CC blood vessels into a tissue, and normally occurs in wound healing,
CC foetal and embryonal development and the formation of the corpus luteum,
CC endometrium and placenta. Angiostatin is a protein (see AAB16450 and
CC AAB68202) involved in angiogenesis, and has an amino acid sequence
CC similar to that of a plasminogen fragment (see murine plasminogen
CC AAB16490). Angiostatin has the ability to inhibit angiogenesis.
CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and
CC AAB68203). Sequences AAB68242 and AAB16522 represent coding and protein
CC sequences of human laminin. Laminin is an angiostatin binding protein,
CC and some of the peptides of the invention share homology with regions of
CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the
CC angiogenesis-inhibiting protein receptor fragments of the invention. The
CC peptides bind either angiostatin or endostatin and can be used in methods
CC for treating diseases and processes that are mediated by angiogenesis,
CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,
CC Crohn's disease, cerebral collaterals, arteriovenous malformations,
CC rubrospinal, diabetic retinopathy, arthritis, wound healing, peptic ulcers,
CC Helicobacter related diseases, fractures, placental and cat scratch
CC fever. They are useful for the detection and prognosis of cancer. DNA
CC sequences A628204-A628241 encode the peptides of the invention.

XX Sequence 183 AA:

SO Query Match 100.0%; Score 893; DB 21; Length 183;

Best Local Similarity 100.0%; Pred. No. 3.9e-101;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLPSRLQDLYSIVRRADRAAPV 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLPSRLQDLYSIVRRADRAAPV 72
QY 61 NLKDELLPSPWEALFSGSGEPLKPGARIFSPDGKDVLNHPMPQKSVHSGSDPNRRRLTE 120
DB 73 NLKDELLPSPWEALFSGSGEPLKPGARIFSPDGKDVLNHPMPQKSVHSGSDPNRRRLTE 132
QY 121 SYCETWRTAPSATGOASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGOASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 182

RESULT 13

AA970771 AAY90771 standard; Protein: 183 AA.

AC AAY90771;

DT 22-AUG-2000 (first entry)

DE Human angiogenesis inhibiting factor 1 protein.

KW Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;

KW abnormal vessel disease.

OS Homo sapiens.

PN CN1244536-A.

PD 16-FEB-2000.

PF 10-AUG-1998; 98CN-0117150.

PR 10-AUG-1998; 98CN-0117150.

XX (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.

XX Yang Z, Guo W;

DR WPI: 2000-388168/34.

DR N-PSDB: AAA29884.

XX Angiogenesis inhibiting factor 1 and its derivative useful for treating

PT tumors -

PS Claim 1; Fig 5; 41pp; Chinese.

XX The present sequence represents an angiogenesis inhibiting factor (I),
CC designated IAF-1. The present invention also describes: (1) preparation
CC of (I) and its derivative; (2) an IAF binding acceptor and its
CC preparation; and (3) an IAF antibody. (I) is useful for preparing new
CC biological preparations for effectively treating various tumours and
CC abnormal-vessel diseases. The IAF antibody is preferably a polyclonal
CC antibody, mosaic antibody, single stranded antibody and human originated
CC antibody.

SO Sequence 183 AA:

Query Match 100.0%; Score 893; DB 21; Length 183;

Best Local Similarity 100.0%; Pred. No. 3.9e-101;

Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLPSRLQDLYSIVRRADRAAPV 60
DB 13 VALNSPLSGMGRGIRGADFOCFQOARAVGLAGTFRAFLPSRLQDLYSIVRRADRAAPV 72

QY 61 NLKDELLPSPWEALFSGSGEPLKPGARIFSPDGKDVLNHPMPQKSVHSGSDPNRRRLTE 120
DB 73 NLKDELLPSPWEALFSGSGEPLKPGARIFSPDGKDVLNHPMPQKSVHSGSDPNRRRLTE 132

QY 121 SYCETWRTAPSATGOASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 170
DB 133 SYCETWRTAPSATGOASSLLGRLGSGAASCHHAYIVLCIENSFMTAS 182

RESULT 14

AA770252 AAY70252 standard; Protein: 183 AA.

AC AAY70252;

DT 06-JUN-2000 (first entry)

DE Human angiogenesis inhibitor, endostatin.

XX Human; immunoglobulin gamma Fc fragment; endostatin; immunofusin;
KW angiogenesis inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;
KW vasotrophic; vulnery; treatment; antiarteriosclerosis; tumour;
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW wound granulation; keloid scar; gene therapy.

OS Homo sapiens.

PN WO200011033-A2.

PD 02-MAR-2000.

PF 25-AUG-1999; 99WO-US19329.

PR 25-AUG-1998; 98US-0097883.

PA (LEXI-) LEXINGEN PHARM CORP.

PI LO K, LI Y, GALLIES SD;

XX WI: 2000-237616/20.
DR N-PSDB: AA251291.
XX
XX Novel fusion protein of angiostatin or endostatin and an immunoglobulin
PT FC region, useful for treating conditions mediated by angiogenesis,
PT such as rheumatoid arthritis, tumors and macular degeneration -
XX
XX Example 1: Pages 41-42; 68pp; English.
XX
XX The patent discloses a DNA molecule encoding a fusion protein comprising
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis
CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment
CC having angiostatin activity, a collagen XVIII fragment having endostatin
CC activity, or combinations of them. The fusion protein (immunofusin) is
CC used to inhibit angiogenesis and to treat diseases or conditions mediated
CC by angiogenesis. Conditions that may be treated include solid tumours,
CC blood born tumours, tumour metastasis, benign tumours including
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental
CC fibroplasia, rubeosis and Osler-Weber Syndrome, myocardial angiogenesis,
CC plaque neovascularisation, telangiectasia, haemophilic joints,
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of
CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used
CC in gene therapy. The present sequence is a human endostatin used in the
CC construction of immunofusin containing human immunoglobulin gamma
CC (IgG) Fc fragment.
XX
SQ Sequence 183 AA:
Query Match 100.0%; Score 893; DB 21; Length 183;
Best Local Similarity 100.0%; Pred. No. 3.9e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VALNSPLSGMGRIGADFOCFQOARAVGLAGTFRAPFLSSRLQDLXSYVRADRAAVPIY 60
DB 13 VALNSPLSGMGRIGADFOCFQOARAVGLAGTFRAPFLSSRLQDLXSYVRADRAAVPIY 72
QY 61 NLKDELLFPPSWAELFSGSEGLPKGARIFFSDGKDVLRHPWPPOKSVHGSDDPNGRRLTE 120
DB 73 NLKDELLFPPSWAELFSGSEGLPKGARIFFSDGKDVLRHPWPPOKSVHGSDDPNGRRLTE 132
QY 121 SYCETWTEAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFWMTAS 170
DB 133 SYCETWTEAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFWMTAS 182
SQ
BUILT 15
ID AAU00896 standard; Protein: 183 AA.
XX AAU00896:
AC
XX
DT 04-JUL-2001 (first entry)
XX
XX Human Endostatin(TM) protein.
DE
XX Human: Endostatin(TM): angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyrogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophilic joint; angiofibroma; wound granulation.
XX
XX Homo sapiens.
OS
XX
XX MO200119989-A2.

XX 22-MAR-2001.
PD
XX 14-SEP-2000; 2000MO-US25166.
PE
XX 14-SEP-1999; 99US-0153698.
PR
XX (ENTR-) ENTREMED INC.
PA
XX Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ,
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL,
PI
XX WI: 2001-244802/25.
DR N-PSDB: AAS00867.
XX
XX Producing Endostatin protein for treating angiogenesis mediated
PT diseases such as solid tumours, comprises recombinantly producing the
PT protein using an expression system, and recovering and purifying the
PT protein -
PS Claim 5: Page 29; 67pp; English.
XX
XX The sequence represents Human Endostatin(TM). The new method of the
CC invention is useful for producing, recovering and purifying Endostatin
CC (TM) from biological sources, such as biological fluids, tissues, cells,
CC culture media, and fermentation media. Endostatin(TM) is useful for
CC treating angiogenesis mediated diseases such as solid tumours, blood
CC borne tumours, leukaemias, tumour metastases, benign tumours, e.g.
CC haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyrogenic
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular
CC degeneration, corneal graft rejection, neovascular glaucoma, colon
CC cancer, retrolental fibroplasia, rubeosis, Osler-Weber Syndrome,
CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,
CC haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM)
CC is also useful for treating disease of excessive or abnormal stimulation
CC of endothelial cells such as intestinal adhesions, atherosclerosis,
CC scleroderma and hypertrophic scars. Higher yields of more purified, and
CC biologically active Endostatin(TM) are obtained by the new method.
CC Endostatin(TM) can be stored in buffers for extended periods of time, and
CC also subjected to lyophilisation, while preserving biological activity.
CC Centrifugation of broth from fermentation steps in production is avoided,
CC preventing unwanted potential cellular lysis and contamination with
CC additional proteins, pigments, enzymes and other cellular chemicals and
CC debris.
XX
SQ Sequence 183 AA:
Query Match 100.0%; Score 893; DB 22; Length 183;
Best Local Similarity 100.0%; Pred. No. 3.9e-101;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VALNSPLSGMGRIGADFOCFQOARAVGLAGTFRAPFLSSRLQDLXSYVRADRAAVPIY 60
DB 13 VALNSPLSGMGRIGADFOCFQOARAVGLAGTFRAPFLSSRLQDLXSYVRADRAAVPIY 72
QY 61 NLKDELLFPPSWAELFSGSEGLPKGARIFFSDGKDVLRHPWPPOKSVHGSDDPNGRRLTE 120
DB 73 NLKDELLFPPSWAELFSGSEGLPKGARIFFSDGKDVLRHPWPPOKSVHGSDDPNGRRLTE 132
QY 121 SYCETWTEAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFWMTAS 170
DB 133 SYCETWTEAPSATGQASSLLGRLGQSAASCHHAYIVLCIENSFWMTAS 182
SQ
Search completed: July 24, 2003, 17:06:04
Job time : 85 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:20:14 ; Search time 52 Seconds
(without alignments)
388.253 Million cell updates/sec

Title: US-09-171-607A-1

Perfect score: 893

Sequence: 1 VANSPLSGMRCIRGADFO.....ASCHHAYIVLCIENSEFWTAS 170

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

al number of hits satisfying chosen parameters: 214442

Minimum DB seq length: 0

Maximum DB seq length: 170

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Published_Applications_AA:*
2: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	38.7	63	9	US-09-822-540A-1
2	182	20.4	31	9	US-09-822-540A-2
3	94	10.5	16	10	US-09-766-412-32
4	63.5	7.1	137	11	US-09-764-872-310
5	63.5	7.1	160	16	US-10-174-693-339
6	63	7.1	91	9	US-09-925-299-1022
7	63	7.1	91	11	US-09-925-299-1022
8	61.5	6.9	92	9	US-09-925-299-805
9	61.5	6.9	92	11	US-09-925-299-805
10	61	6.8	104	11	US-09-764-891-5225
11	61	6.8	133	15	US-10-023-282-534
12	60.5	6.8	64	15	US-10-023-282-1109
13	60.5	6.8	128	11	US-09-764-891-2789
14	60.5	6.8	128	15	US-10-205-428-262
15	60	6.7	135	14	US-10-014-717-23

16	59	6.6	14	10	US-09-766-412-31	Sequence 31, App1
17	59	6.6	116	9	US-09-867-550-1976	Sequence 1976, Ap
18	59	6.6	162	15	US-10-156-761-11736	Sequence 11736, A
19	58.5	6.6	133	15	US-10-198-070-45	Sequence 45, App1
20	58	6.5	64	11	US-09-866-050A-122	Sequence 122, App
21	57.5	6.4	151	10	US-09-764-864-1009	Sequence 1009, Ap
22	57	6.4	116	10	US-09-796-692-1590	Sequence 1590, Ap
23	57	6.4	116	15	US-10-040-862-1590	Sequence 1590, Ap
24	57	6.4	157	10	US-09-764-868-913	Sequence 913, App
25	57	6.4	169	9	US-09-925-297-500	Sequence 500, App
26	56.5	6.3	109	9	US-09-864-761-34592	Sequence 34592, A
27	56.5	6.3	113	9	US-09-864-761-41366	Sequence 41366, A
28	56.5	6.3	141	15	US-10-106-698-4881	Sequence 4881, Ap
29	56.5	6.3	162	9	US-09-727-801-4	Sequence 4, App1
30	56.5	6.3	169	9	US-09-927-602-9	Sequence 9, App1
31	56	6.3	88	10	US-09-764-877-1924	Sequence 1924, Ap
32	56	6.3	165	9	US-09-842-528-4	Sequence 4, App1
33	55.5	6.2	79	15	US-10-157-031-204	Sequence 204, App
34	55.5	6.2	81	15	US-10-106-698-5181	Sequence 5181, App
35	55	6.2	86	10	US-09-987-108-12	Sequence 12, App1
36	55	6.2	121	9	US-09-864-761-40151	Sequence 40151, A
37	54.5	6.1	170	15	US-10-026-741-65	Sequence 65, App1
38	54	6.0	86	10	US-09-987-108-13	Sequence 13, App1
39	54	6.0	96	11	US-09-764-881-161	Sequence 161, App
40	54	6.0	138	15	US-10-106-698-6854	Sequence 6854, App
41	54	6.0	146	10	US-09-479-040-5	Sequence 5, App1
42	54	6.0	151	9	US-09-925-301-1144	Sequence 1144, App
43	53.5	6.0	134	15	US-10-102-806-596	Sequence 596, App
44	53.5	6.0	154	15	US-10-106-698-5838	Sequence 5838, Ap
45	53.5	6.0	168	11	US-09-862-540-46	Sequence 46, App1

ALIGNMENTS

US-09-822-540A-1	US-09-822-540A-1
Sequence 1, Application US/09822540A	
Patent No. US20020058620A1	
GENERAL INFORMATION:	
APPLICANT: Phoenix Pharmaceuticals, Inc.	
APPLICANT: Chang, Jaw-Kang	
TITLE OF INVENTION: Cell Growth Regulation System	
FILE REFERENCE: PhoenixPharFull	
CURRENT APPLICATION NUMBER: US/09/822,540A	
CURRENT FILING DATE: 2001-03-30	
PRIOR APPLICATION NUMBER: 60/194,561	
PRIOR FILING DATE: 2000-04-03	
NUMBER OF SEQ ID NOS: 2	
SOFTWARE: PatentIn version 3.1	
SEQ ID NO 1	
LENGTH: 63	
TYPE: PRT	
ORGANISM: Homo sapiens	
US-09-822-540A-1	
Query Match	38.7% Score 346; DB 9; Length 63;
Best Local Similarity	100.0%; Pred. No. 2, 2e-31;
Matches 63; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	52 ADRAAYPIVINDDELFPSEWALFSSSEGLPRGATIFSFDDKDVLRHPTWPKSWYHGS 111
DB	1 ADRAAYPIVINDDELFPSEWALFSSSEGLPRGATIFSFDDKDVLRHPTWPKSWYHGS 60
OY	112 DPN 114
DB	61 DPN 63
RESULT 2	
US-09-822-540A-2	
Sequence 2, Application US/09822540A	
Patent No. US20020058620A1	

```

: RESULT 4
: US-09-764-872-310
: Sequence 310, Application US/09764872
: Publication No. US20030050231A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PA125
: CURRENT APPLICATION NUMBER: US/09/764,872
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 957
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 310
: LENGTH: 137
: TYPE: PRT

```

[illegible]

Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1022
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (39)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
-09-925-299-1022

Query Match
Best Local Similarity 7.1%; Score 63; DB 9; Length 91;
Matches 25; Conservative 11; Mismatches 29; Indels 22; Gaps 6;
OY 81 PLKPGARIEFDGKDVLRHPMPQKSVHSGDPNGRRLTESYCEFTWTEAPS--ATG-- 135
DB 18 PLSPGLRV-----VYGH-TWRFVYVFXTE-----FHSCCPGMSAMAPSRLLTATSTS 63
OY 136 --QASSLLGRLGGSASACHHAYIVL 160
DB 64 WFKRSQASASQVVGITGA-CHHTWTLIL 89

RESULT 7
US-09-925-299-1022
Sequence 1022, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1022
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (39)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1022

Query Match
Best Local Similarity 7.1%; Score 63; DB 11; Length 91;
Matches 25; Conservative 11; Mismatches 29; Indels 22; Gaps 6;
OY 81 PLKPGARIEFDGKDVLRHPMPQKSVHSGDPNGRRLTESYCEFTWTEAPS--ATG-- 135
DB 18 PLSPGLRV-----VYGH-TWRFVYVFXTE-----FHSCCPGMSAMAPSRLLTATSTS 63
OY 136 --QASSLLGRLGGSASACHHAYIVL 160
DB 64 WFKRSQASASQVVGITGA-CHHTWTLIL 89

RESULT 8
US-09-925-299-805
Sequence 805, Application US/09925299
Publication No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 805
LENGTH: 92
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-299-805

Query Match
Best Local Similarity 6.9%; Score 61.5; DB 9; Length 92;
Matches 19; Conservative 7; Mismatches 31; Indels 13; Gaps 2;
OY 101 TWPQKSVHSGDPNGRRLTESYCEFTWTEAPSATGQASSLLGRLGGSASACHHAYIVL 160
DB 8 TWPKRAEMFAGAPGQ-----WPRSLSYVHSTTRC---PLVGVRABGLRHAYIVL 54
OY 161 CIENSFWTAS 170
DB 55 ELGTTDWTGS 64

RESULT 9
US-09-925-299-805
Sequence 805, Application US/09925299
Publication No. US20030040617A9
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 805
LENGTH: 92
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-299-805

Query Match
Best Local Similarity 6.9%; Score 61.5; DB 11; Length 92;
Matches 19; Conservative 7; Mismatches 31; Indels 13; Gaps 2;
OY 101 TWPQKSVHSGDPNGRRLTESYCEFTWTEAPSATGQASSLLGRLGGSASACHHAYIVL 160
DB 8 TWPKRAEMFAGAPGQ-----WPRSLSYVHSTTRC---PLVGVRABGLRHAYIVL 54
OY 161 CIENSFWTAS 170
DB 55 ELGTTDWTGS 64

RESULT 10

Query Match	6.88;	Score 61;	DB 11;	Length 104;
Best Local Similarity	28.9%;	Pred. NO. 35;		
Matches 24;	Conservative 8;	Mismatches 29;	Indels 22;	Gaps 3

```

: Sequence 534, Application US/10023282
: Publication No. US20030092893A1
: GENERAL INFORMATION:
: APPLICANT: Young et al.
: TITLE OF INVENTION: 207 Human Secreted Proteins
: Filing Reference: 0700701

```

```
FILE REFERENCE: P200/P1
CURRENT APPLICATION NUMBER: US/10/023,282
CURRENT FILING DATE: 2001-12-20
EARLIER APPLICATION NUMBER: 09/205,258
EARLIER FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US99/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
```

Query Match	6.88;	Score 61;	DB 15;	Length 133;
Best Local Similarity	38.18;	Pred. NO. 48;		
Matches 16; Conservative	6;	Mismatches 20;	Indels 0;	Gaps 0;

RESULT 12
US-10-023-282-1109

12

OY 156 AYVLCI 162
DB 101 RYVYICL 107

RESULT 14

US-10-205-428-262
; Sequence 262, Application US/10205428
; Publication No. US20030108907A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P117C1
; CURRENT APPLICATION NUMBER: US/10/205,428
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1019
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 262
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-428-262

Query Match 6.8%; Score 60.5; DB 15; Length 128;
Best Local Similarity 24.4%; Pred. No. 52;
Matches 31; Conservative 12; Mismatches 39; Indels 45; Gaps 7;

DB 56 AVPIVNLDELLEPSEWAL-----FSGSEGPLKPGARIFSPDGKDYLRHPT 101
6 SVSTFTLSPETGFPWPSVPIPLPTAQWLHRAWSSKEDGLK-----SDICEEP- 55
OY 102 WPKSVWHSNDPNGRRLTES--YCETWRTAPSATGASSLLGGRLLGQSAASC---HH 155
DB 56 -----SCSKSHSPSGDLMLFSLIYC-----SYNPEALDLFGG---SQKGISCLPIINQ 100
OY 156 AYVLCI 162
DB 101 RYVYICL 107

RESULT 15

US-10-014-717-23
; Sequence 23, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHIOLONES

; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 23
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-10-014-717-23

Query Match. 6.7%; Score 60; DB 14; Length 135;
Best Local Similarity 31.5%; Pred. No. 63;
Matches 23; Conservative 7; Mismatches 27; Indels 16; Gaps 3;

OY 76 SGSEGP---LKPGRIFSPDGKDYLRHPTWPKSVWHSNDPNGRRLTESYCETWRTAP 131
DB 22 SSGACARAGRAHGAASAGFEGGDYMR-----KARAHGAMLGGRD-----DGMRRGLP 69
OY 132 SATGQASSLLGR 144
DB 70 GAGALRALALORGR 82

Search completed: July 24, 2003, 17:30:07
Job time : 53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2003, 17:17:23 ; Search time 29 Seconds
(without alignments)
248.029 Million cell updates/sec

Title: US-09-171-607a-1

Perfect score: 893
Sequence: 1 VALNSPLSGMGRGIRGADFDQ.....ASCHHAYIVLCIENSFMFTAS 170

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

254538

Minimum DB seq length: 0
Maximum DB seq length: 170

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160	17.9	35	3	US-09-046-985-2
2	160	17.9	35	3	US-09-474-743-2
3	101	11.3	22	3	US-09-046-985-7
4	101	11.3	22	3	US-09-474-743-7
5	94	10.5	16	3	US-09-385-442-32
6	66	7.4	22	3	US-09-046-985-10
7	66	7.4	22	3	US-09-474-743-10
8	66	7.4	22	3	US-09-046-985-9
9	66	7.4	22	3	US-09-474-743-9
10	66	7.4	22	3	US-09-046-985-12
11	66	7.4	26	3	US-09-474-743-12
12	65.5	7.3	128	4	US-09-252-991A-20717
13	64	7.2	20	3	US-09-046-985-11
14	64	7.2	20	3	US-09-474-743-11
15	63.5	7.1	87	2	US-08-461-9908-20
16	63.5	7.1	138	4	US-09-252-991A-27071
17	63.5	7.1	160	4	US-09-615-192A-339
18	62	6.9	170	4	US-09-252-991A-18663
19	61.5	6.9	92	4	US-09-673-809-106
20	61.5	6.9	135	4	US-09-252-991A-31675
21	61.5	6.9	153	4	US-09-252-991A-30249
22	61	6.8	22	3	US-09-046-985-4
23	61	6.8	22	3	US-09-474-743-4
24	61	6.8	133	4	US-09-205-258-534
25	61	6.8	151	4	US-09-252-991A-19354
26	60.5	6.8	64	4	US-09-205-258-1109
27	60	6.7	135	3	US-09-335-409-23

28	60	6.7	135	4	US-09-568-102-23	Sequence 23, Appl
29	60	6.7	135	4	US-09-567-969-23	Sequence 23, Appl
30	60	6.7	135	4	US-09-568-480-23	Sequence 23, Appl
31	60	6.7	135	4	US-09-568-486-23	Sequence 23, Appl
32	60	6.7	135	4	US-09-568-472-23	Sequence 23, Appl
33	60	6.7	135	4	US-09-567-899-23	Sequence 23, Appl
34	59	6.6	14	3	US-09-385-442-31	Sequence 31, Appl
35	58.5	6.6	141	4	US-09-252-991A-26729	Sequence 26729, A
36	58	6.5	64	3	US-09-188-930-122	Sequence 122, App
37	57.5	6.4	97	3	US-08-905-223-444	Sequence 444, App
38	57.5	6.4	124	4	US-09-252-991A-28290	Sequence 28290, A
39	57.5	6.4	97	3	US-08-905-223-444	Sequence 20099, A
40	57.5	6.4	155	4	US-09-252-991A-20099	Sequence 1, Appl
41	57	6.4	113	1	US-08-455-645D-1	Sequence 1, Appl
42	57	6.4	113	5	PCr-US96-07925-1	Sequence 25930, A
43	57	6.4	146	4	US-09-252-991A-25930	Sequence 22731, A
44	57	6.4	169	4	US-09-252-991A-22731	Sequence 20369, A
45	56.5	6.3	133	4	US-09-252-991A-20369	

ALIGNMENTS

Result 1
US-09-046-985-2
Sequence 2, Application US/09046985
Patent No. 6121236
GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE ANGIOGENESIS
TITLE OF INVENTION: ANGIOGENESIS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,985
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CMCC-614
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-046-985-2
Query Match 17.9% Score 160; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QARAVGLAGTFRALSSRLQDLXSYVRADRAV 57
|||||
Db 1 QARAVGLAGTFRALSSRLQDLXSYVRADRAV 34
|||||

```
RESULT 2
US-09-474-743-2
; Sequence 2, Application US/09474743
; Patent No. 6235716
; GENERAL INFORMATION:
; APPLICANT: Ben-Sasson, Shmuel A.
; TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
;   TITLE OF INVENTION: ANGIOGENESIS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/474,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/046,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-474-743-2

Query Match      17.9%; Score 160; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.4e-12;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      24 QARAVGLAGTRAFLLSRLODLYSIVRRADRAAV 57
      |||||||||||||||||||||||||||||||||||
      1 QARAVGLAGTRAFLLSRLODLYSIVRRADRAAV 34

RESULT 3
US-09-046-985-7
; Sequence 7, Application US/09046985
; Patent No. 612136
; GENERAL INFORMATION:
; APPLICANT: Ben-Sasson, Shmuel A.
; TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
;   TITLE OF INVENTION: ANGIOGENESIS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/474,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/046,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= modified aa
; OTHER INFORMATION: /note= "N-Acetyl-L-threonine"
```

```
SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,985
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= modified aa
; OTHER INFORMATION: /note= "N-Acetyl-L-threonine"
US-09-046-985-7

Query Match      11.3%; Score 101; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

      0Y      33 TFRFLSSRLQDLYSIVRRAD 53
      |||||||||||||||||||||||||||||||
      DB      1 TFRFLSSRLQDLYSIVRRAD 21

RESULT 4
US-09-474-743-7
; Sequence 7, Application US/09474743
; Patent No. 6235716
; GENERAL INFORMATION:
; APPLICANT: Ben-Sasson, Shmuel A.
; TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
;   TITLE OF INVENTION: ANGIOGENESIS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/474,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/046,985
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
```

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "N-Acetyl-Threonine"
US-09-474-743-7

Query Match 11.3%; Score 101; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.2e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 33 TFFAFLSSRLQDLYSIVRRAD 53
1 TFFAFLSSRLQDLYSIVRRAD 21

DB 1 TFFAFLSSRLQDLYSIVRRAD 21

SUPT 5
-09-385-442-32
Sequence 32, Application US/09385442
Patent No. 6200954
GENERAL INFORMATION:
APPLICANT: Ge, Ruowen
APPLICANT: Kint, R. Manjunatha
TITLE OF INVENTION: Small Peptides Having Potent Anti-Angiogenic Activity
FILE REFERENCE: 1781-170P
CURRENT APPLICATION NUMBER: US/09/385,442
EARLIER FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 60/099,313
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 32
LENGTH: 16
TYPE: PRT
ORGANISM: mammalian
FEATURE:
OTHER INFORMATION: Endo-4
US-09-385-442-32

Query Match 10.5%; Score 94; DB 3; Length 16;
Best Local Similarity 87.5%; Pred. No. 5.4e-05;
Matches 14; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 99 HPTWPKSVWVGSDPN 114
1 HPTWPKSVWVGSDPS 16

RESULT 6
US-09-046-985-10
Sequence 10, Application US/09046985
Patent No. 6121236
GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,985
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CMCC-614
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "N-Acetyl-Alanine"
US-09-046-985-10

Query Match 7.4%; Score 66; DB 3; Length 22;
Best Local Similarity 71.4%; Pred. No. 0.23;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 34 FRAFLSSRLQDLYSIVRRADR 54
2 FRAFLSSRLQDLYSIVRRADR 22

DB 2 FRAFLSSRLQDLYSIVRRADR 22

RESULT 7
US-09-474-743-10
Sequence 10, Application US/09474743
Patent No. 6235716
GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/474,743
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/046,985
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CMCC-614
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid

STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "N-Acetyl-Alanine"
US-09-474-743-10

Query Match 7.4% Score 66; DB 3; Length 22;
Best Local Similarity 71.4%; Pred. No. 0.23;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 34 FRAFLSSRLQDLYSIVRRADR 54
: | | | :|||||
Db 2 YRWRLSHRPKDLYSIVRRADR 22

RESULT 8

US-09-046-985-9
Sequence 9, Application US/09046985
Patent No. 6121236

GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,985
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "N-Acetyl-Threonine"
US-09-046-985-9

Query Match 7.4% Score 66; DB 3; Length 23;
Best Local Similarity 71.4%; Pred. No. 0.24;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 34 FRAFLSSRLQDLYSIVRRADR 54
: | | | :|||||
Db 3 YRWRLSHRPKDLYSIVRRADR 23

RESULT 9

US-09-474-743-9
Sequence 9, Application US/09474743
Patent No. 6235716

GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/474,743
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/046,985
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CMCC-614
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "N-Acetyl-Threonine"
US-09-474-743-9

Query Match 7.4% Score 66; DB 3; Length 23;
Best Local Similarity 71.4%; Pred. No. 0.24;
Matches 15; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 34 FRAFLSSRLQDLYSIVRRADR 54
: | | | :|||||
Db 3 YRWRLSHRPKDLYSIVRRADR 23

RESULT 10

US-09-046-985-12
Sequence 12, Application US/09046985
Patent No. 6121236

GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts

RESULT 11
 US-09-474-743-12
 Sequence 12, Application US/09474743
 Patent No. 6235716
 GENERAL INFORMATION:
 APPLICANT: Ben-Sasson, Shmuel A.
 TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
 TITLE OF INVENTION: ANGIOGENESIS
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 STREET: Two Millitta Drive
 CITY: Lexington
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02173
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/474,743
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/046,985
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brook, David E.
 REGISTRATION NUMBER: 22,592
 REFERENCE/DOCKET NUMBER: CMCC-614

RESULT 13
 US-09-046-985-11
 ; Sequence 11, Application US/09046985
 ; Patent No. 6121236
 ; GENERAL INFORMATION:
 ; APPLICANT: Ben-Sasson, Shmuel A.
 ; TITLE OF INVENTION: NOVEL, MULTIVALENT LIGANDS WHICH MODULATE
 ; TITLE OF INVENTION: ANGIOGENESIS
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Millitia Drive

CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/046,985
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CMCC-614
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO.: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "Alpha N-Acetyl Arginine"
US-09-046-985-11

Query Match 7.2%; Score 64; DB 3; Length 20;
Best Local Similarity 82.4%; Pred. No. 0.35;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 38 LSSRLQDLYSYVRADR 54
DB 4 LSHRPKDLXSYVRADR 20

RESULT 14
US-09-474-743-11
Sequence 11, Application US/09474743
Patent No. 6235716
GENERAL INFORMATION:
APPLICANT: Ben-Sasson, Shmuel A.
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE
TITLE OF INVENTION: ANGIOGENESIS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/474,743
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/046,985
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.

REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CMCC-614
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO.: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= modified aa
OTHER INFORMATION: /note= "Alpha N-Acetyl Arginine"
US-09-474-743-11

Query Match 7.2%; Score 64; DB 3; Length 20;
Best Local Similarity 82.4%; Pred. No. 0.35;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 38 LSSRLQDLYSYVRADR 54
DB 4 LSHRPKDLXSYVRADR 20

RESULT 15
US-08-461-990B-20
Sequence 20, Application US/08461990B
Patent No. 5851810
GENERAL INFORMATION:
APPLICANT: JOHN S. BLANCHARD
TITLE OF INVENTION: NUCLEIC ACID ENCODING RHODOCOCCUS
TITLE OF INVENTION: PHENYLALANINE DEHYDROGENASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,990B
FILING DATE: JUNE 5, 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG J. ARNOLD
REGISTRATION NUMBER: 34,287
REFERENCE/DOCKET NUMBER: 96700/370
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO.: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 87
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
DESCRIPTION: NO
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: C. SYMBIOSUM
INDIVIDUAL ISOLATE: GLUTAMATE DEHYDROGENASE
US-08-461-990B-20

Query Match 7.1%; Score 63.5; DB 2; Length 87;
 Best Local Similarity 34.7%; Pred. No. 3.3;
 Matches 25; Conservative 6; Mismatches 22; Indels 19; Gaps 5;

OY	75	ESGSEGLKPGAR-----IFSDG-----KDYLRHPTWPKSVWHGS--DPNGR--R	117
Db	3	FNGAIGPYKGLRFPAPSVNLISIMKFLGFEOAFKDSL--TTLPMGCAKCGSDDPNGKSDR	60
OY	118	LTESYCETWTE	129
Db	61	EVMRFCAFWTE	72

Search completed: July 24, 2003, 17:22:23
 Job time : 31 secs

THIS PAGE BLANK (USPTO)